

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:53:36 ; Search time 508.584 Seconds
(without alignments)
1960.127 Million cell updates/sec

Title: US-09-889-611a-19

Perfect score: 23

Sequence: 1 gaatgaactacatacaaacacc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.ov.*

5: gb.pat.*

6: gb.ph.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.sts.*

11: gb.sy.*

12: gb.un.*

13: gb.vi.*

14: gb.ba.*

15: em.fun.*

16: em.fun.*

17: em.fun.*

18: em.in.*

19: em.mu.*

20: em.mu.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rtd.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	23	100.0	4229	9	AF234618	AF234618 Homo sapi
2	23	100.0	157284	9	AC072051	AC072051 Homo sapi
3	23	100.0	189092	2	AC027524	AC027524 Homo sapi
4	20.4	88.7	6182	6	AX346917	AX346917 Sequence
5	18.4	80.0	687	11	BV074238	BV074238 S212P6017
6	18.4	80.0	57941	9	BX323827	BX323827 Human DNA
7	18.4	80.0	110000	2	AL928982	Continuation (5 of
8	18.4	80.0	114390	8	AP005113	AP005113 Oryza sat
9	18.4	80.0	119420	9	HS997K18	AL021406 Human DNA
10	18.4	80.0	207166	2	AC073703	AC073703 Mus muscu
11	18.4	80.0	316678	2	AC132190	AC132190 Mus muscu
12	18.2	79.1	124786	9	AC115285	AC115285 Homo sapi
13	18.2	79.1	157653	9	AC005146	AC005146 Homo sapi
14	18.2	79.1	202398	2	AC025303	AC025303 Homo sapi
15	18.2	79.1	238554	2	AC020965	AC020965 Mus muscu
16	18	78.3	37490	9	AC026356	AC026356 Homo sapi
17	18	78.3	156394	2	AC026558	AC026558 Homo sapi
18	18	78.3	215960	2	AC125959	AC125959 Rattus no
19	18	78.3	220480	2	AC023973	AC023973 Homo sapi
20	17.8	77.4	31199	2	AC109976	AC109976 Rattus no
21	17.8	77.4	105308	8	AF273333	AF273333 Lycopersi
22	17.8	77.4	125290	2	AC145820	AC145820 Cicer ari
23	17.8	77.4	125990	9	AC108734	AC108734 Homo sapi
24	17.8	77.4	131209	9	AC126177	AC126177 Homo sapi
25	17.8	77.4	149806	10	AC105159	AC105159 Mus muscu
26	17.8	77.4	152898	2	AC134432	AC134432 Mus muscu
27	17.8	77.4	162446	9	AL137118	AL137118 Human DNA
28	17.8	77.4	163012	2	AC142104	AC142104 Mus muscu
29	17.8	77.4	164603	2	AC119466	AC119466 Rattus no
30	17.8	77.4	167170	10	AL669968	AL669968 Mouse DNA
31	17.8	77.4	167229	2	AC101950	AC101950 Mus muscu
32	17.8	77.4	176278	10	AC126557	AC126557 Mus muscu
33	17.8	77.4	181060	5	BX119907	BX119907 Zebrafish
34	17.8	77.4	182116	9	AC093831	AC093831 Homo sapi
35	17.8	77.4	185211	8	OSJN00158	AL662955 Oryza sat
36	17.8	77.4	187532	9	AC006333	AC006333 Homo sapi
37	17.8	77.4	192092	10	AC126248	AC126248 Mus muscu
38	17.8	77.4	193856	2	BX537353	BX537353 Danio rer
39	17.8	77.4	194778	2	AC118844	AC118844 Rattus no
40	17.8	77.4	196265	10	AC116325	AC116325 Mus muscu
41	17.8	77.4	198338	10	AC087098	AC087098 Genomic s
42	17.8	77.4	211236	2	AC120182	AC120182 Mus muscu
43	17.8	77.4	220320	2	AC132527	AC132527 Rattus no
44	17.8	77.4	228428	10	AL596456	AL596456 Mouse DNA

ALIGNMENTS

RESULT 1
AF234618 4229 bp DNA linear PRI 29-MAR-2002
LOCUS Homo sapiens meggin gene, promoter region and partial sequence.
DEFINITION AF234618
ACCESSION AF234618.2 GI:19808130
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 4229)
Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
Transcriptional regulation of a mesangium-predominant gene, meggin
Unpublished


```

/clone_lib="RPC1-11 Human Male BAC"
599. .635
/rpt_family="(T)n"
1040. .1077
/rpt_family="AT_rich"
1411. .1747
/rpt_family="LIP"
2632. .2992
/rpt_family="MLT1A1"
3115. .3138
/rpt_family="AT_rich"
30596. .30624
/rpt_family="LIMC4a"
3834. .4338
/rpt_family="MLT1C"
complement(4339. .5397)
/rpt_family="LIMC4a"
complement(5742. .5899)
/rpt_family="MER5B"
6499. .6816
/rpt_family="HAL1"
complement(7416. .7621)
/rpt_family="MER3"
complement(8163. .9090)
/rpt_family="L2"
9276. .9419
/rpt_family="MIR"
9598. .9658
/rpt_family="MER81"
9659. .10033
/rpt_family="MLT1J"
13985. .14009
/rpt_family="AT_rich"
14011. .14115
/rpt_family="(TA)n"
complement(14942. .15044)
/rpt_family="MIR3"
15155. .15298
/rpt_family="L2"
15592. .15773
/rpt_family="LIM4c"
15874. .18310
/rpt_family="LIP1A16"
16373. .16374
/note="1328 bp of bacterial transposon insertion in BAC
excised between these 2 bp"
18311. .18331
/rpt_family="(TAAAA)n"
18332. .19566
/rpt_family="LIP1A16"
19567. .19587
/rpt_family="(CA)n"
19588. .20042
/rpt_family="LIP1A16"
20043. .20064
/rpt_family="AT_rich"
complement(21259. .21386)
/rpt_family="MIR3"
22395. .22420
/rpt_family="AT_rich"
complement(22788. .22948)
/rpt_family="FRAM"
complement(23843. .24253)
/rpt_family="MSTA"
complement(24678. .25041)
/rpt_family="MLT1A2"
25054. .25094
/rpt_family="HAL1"
25214. .25358
/rpt_family="LIM4c"
complement(25602. .25854)
/rpt_family="MIR"
26745. .26796

```

```

repeat_region /rpt_family="AT_rich"
27451. .27484
/rpt_family="(CAAAA)n"
complement(28183. .28219)
/rpt_family="MIR"
complement(28249. .28325)
/rpt_family="MIR"
complement(29372. .29459)
/rpt_family="MIR"
29539. .30003
/rpt_family="MER66B"
30596. .30624
/rpt_family="AT_rich"
31415. .31829
/rpt_family="L2"
31846. .31915
/rpt_family="(TGAA)n"
complement(32019. .32103)

Query Match 100.0%; Score 23; DB 9; Length 157284;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATACACACCACC 23
|||||
DB 155067 GAATGAACCTACATACACACCACC 155045

RESULT 3
AC027524 189092 bp DNA linear HTG 27-MAR-2003
LOCUS Homo sapiens chromosome 18 clone RP11-317G1 map 18, 3 unordered
DEFINITION pieces.
AC027524
VERSION AC027524.4 GI:21307437
KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-317G1
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lechoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Piserre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

```

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, J., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McKernan, K., Meldrum, J., Menees, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, K., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2002 this sequence version replaced gi:11990731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WITBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8516

Center clone name: 317_G_1

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 85659: contig of 85659 bp in length
* 85660 85759: gap of 100 bp
* 85760 163014: contig of 77255 bp in length
* 163015 163114: gap of 100 bp
* 163115 189092: contig of 25978 bp in length.

FEATURES

source
1. .189092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-317G1"
/clone_lib="RPC1-11 Human Male BAC"

ORIGIN

Query Match 100.0%; Score 23; DB 2; Length 189092;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATGAACATACATACACACC 23

|||||

Db 66563 GAATGAACATACATACACACC 66585

RESULT 4

AX346917/c
LOCUS
DEFINITION Sequence 1988 from Patent WO0200928.
ACCESSION AX346917

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .6182

/organism="synthetic construct"

/mol_type="unassigned DNA"

/db_xref="taxon:32630"

/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 88.7%; Score 20.4; DB 6; Length 6182;

Best Local Similarity 95.5%; Pred. No. 55;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AATGAACATACATACACACC 23

|||||

Db 1310 AATGAACATACATACACACC 1289

|||||

RESULT 5

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1. .687

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="CZECHII/E1"

/db_xref="taxon:10090"

/map="12 22-639 20493381-20492764"

FEATURES

source

1. .687

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="CZECHII/E1"

/db_xref="taxon:10090"

/map="12 22-639 20493381-20492764"

WGS-discovery: Paired-end low-coverage whole genome shotgun reads

were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS

reads were placed uniquely on the MGSCV3 C57BL/6J assembly and SNP

detection was carried out by SSAHA-SNP. 225,000 reads were

annotated

as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J

and the strain from which the particular read came. The validation

rate for these SNPs was estimated at approximately 98%.

Location/Qualifiers

1. .687

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="CZECHII/E1"

/db_xref="taxon:10090"

/map="12 22-639 20493381-20492764"

WGS-discovery: Paired-end low-coverage whole genome shotgun reads

were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS

reads were placed uniquely on the MGSCV3 C57BL/6J assembly and SNP

detection was carried out by SSAHA-SNP. 225,000 reads were

annotated

as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J

and the strain from which the particular read came. The validation

rate for these SNPs was estimated at approximately 98%.

Location/Qualifiers

1. .687

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="CZECHII/E1"

/db_xref="taxon:10090"

/map="12 22-639 20493381-20492764"


```

Best Local Similarity 95.0%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AATGAACCTACATAACCA 21
   |||||
Db 49240 AATGAACCTACATAACCA 49259

RESULT 9
HS997K18/c
LOCUS
DEFINITION
  HS997K18 119420 bp DNA linear PRI 10-MAR-2001
  Human DNA sequence from clone RP5-997K18 on chromosome 20p12.
  Contains STSs and GSSs, complete sequence.
ACCESSION
AL021406
VERSION
AL021406.1 GI:2864605
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (Bases 1 to 119420)
Swann,M.
Direct Submission
Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 11, 1998 this sequence version replaced gi:2808693.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP5-997K18 is from the library RPCI-5 constructed by the group of
Piet de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-997K18.
FEATURES
Location/Qualifiers
1..119420
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704K18997"
/db_xref="taxon:9606"
/chromosome="20"
/map="p12"
/clone="RP5-997K18"
/clone_lib="RPCI-5"
3694..3873
/note="3 copies 60 mer 75% conserved"
repeat_region
4066..4401
/note="LMC/D repeat: matches 5186. 5568 of consensus"
repeat_region
5737..5997
/note="LMC/D repeat: matches 5186. 5568 of consensus"
repeat_region
6161..6379
/note="LMC/D repeat: matches 5186. 5568 of consensus"
repeat_region
7895..7963
/note="LMC/D repeat: matches 5186. 5568 of consensus"
repeat_region

/note="TH1B repeat: matches 295. 364 of consensus"
7960..8026
/note="TH1A repeat: matches 1. 41 of consensus"
8027..9238
/note="TH1A-internal repeat: matches 371. 1580 of
consensus"
9239..9549
/note="AluSP repeat: matches 1. 311 of consensus"
9550..9920
/note="TH1A-internal repeat: matches 1. 371 of consensus"
9921..10262
/note="TH1B repeat: matches 1. 364 of consensus"
complement(10771..10852)
/note="match: GSS: Em:AQ187931"
10771..10825
/note="match: GSS: Em:AQ316710"
10876..11057
/note="MIR repeat: matches 17. 189 of consensus"
11197..11368
/note="MER5B repeat: matches 13. 172 of consensus"
12580..12789
/note="LTR16C repeat: matches 177. 382 of consensus"
13911..14092
/note="MER3 repeat: matches 1. 209 of consensus"
14844..14968
/note="MIR repeat: matches 93. 230 of consensus"
15462..15997
/note="L2 repeat: matches 963. 1533 of consensus"
16117..16279
/note="L2 repeat: matches 1767. 1940 of consensus"
16528..16721
/note="MER63A repeat: matches 1. 206 of consensus"
16725..16825
/note="L2 repeat: matches 2647. 2748 of consensus"
17301..17700
/note="MIR repeat: matches 136. 544 of consensus"
17706..17785
/note="BUP1 repeat: matches 12932. 13007 of consensus"
18005..18024
/note="MIR repeat: matches 21. 262 of consensus"
19465..19902
/note="match: GSS: Em:B76059"
20736..20819
/note="3 copies 28 mer 79% conserved"
20989..21238
/note="MIR repeat: matches 3. 262 of consensus"
21636..22204
/note="L2 repeat: matches 2122. 2710 of consensus"
22244..22380
/note="LMC/D repeat: matches 32. 167 of consensus"
23059..23928
/note="LMC/D repeat: matches 974. 1891 of consensus"
24300..24608
/note="AluJo repeat: matches 1. 289 of consensus"
24609..24655
/note="WAD1 repeat: matches 1. 51 of consensus"
24657..24802
/note="73 copies 2 mer aa 59% conserved"
24853..26361
/note="LMC/D repeat: matches 4815. 6326 of consensus"
26636..26978
/note="MIR repeat: matches 14. 364 of consensus"
27148..27183
/note="L2 copies 3 mer tat 86% conserved"
27228..29379
/note="LMC/D repeat: matches 3948. 6153 of consensus"
29380..29690
/note="AluYa5 repeat: matches 1. 311 of consensus"
29691..30066
/note="LMC/D repeat: matches 3572. 3948 of consensus"
30472..30525
/note="MIR repeat: matches 2 mer gt 94% conserved"
30540..30811

```

```

/note="AluX repeat: matches 39. .312 of consensus"
31226. .31518
/note="AluJo repeat: matches 1. .299 of consensus"
32108. .32413
/note="AluX repeat: matches 1. .306 of consensus"
32743. .32894
/note="MIR repeat: matches 50. .213 of consensus"
33424. .33614
/note="match: GSS: Em:B91428"
34293. .34333
/note="match: STS: Em:G09780"
complement(34293. 34333)
/note="match: STS: Em:G09898"
34303. .34332
/note="10 copies 3 mer tta 100% conserved"
34334. .34676
/note="LIP47 repeat: matches 5807. .6145 of consensus"
34716. .34947
/note="LINE repeat: matches 5576. .5810 of consensus"
35064. .35135
/note="18 copies 4 mer tata 69% conserved"
35065. .35132
/note="34 copies 2 mer at 67% conserved"
36181. .36570
/note="LIR16B repeat: matches 38. .458 of consensus"
36610. .36986
/note="L2 repeat: matches 1596. .2008 of consensus"
37799. .37860
/note="MLT1I repeat: matches 74. .135 of consensus"
38576. .38734
/note="LIM49 repeat: matches 5855. .5989 of consensus"
38802. .38913
/note="LIM49 repeat: matches 6163. .6270 of consensus"
40678. .40811
/note="LIM2 repeat: matches 6175. .6308 of consensus"
40812. .41083
/note="LIMC4 repeat: matches 7710. .7970 of consensus"
42665. .42718
/note="LIP3 repeat: matches 5501. .5554 of consensus"
43334. .43362
/note="MLT1J repeat: matches 112. .140 of consensus"
44826. .45020
/note="MLT1A1 repeat: matches 37. .233 of consensus"
45035. .45382
/note="LIMD3 repeat: matches 7330. .7676 of consensus"
45383. .45737
/note="THE1C repeat: matches 1. .371 of consensus"
45738. .45796
/note="LIMD3 repeat: matches 7272. .7330 of consensus"
46215. .47939
/note="LIP45 repeat: matches 4407. .6141 of consensus"
48417. .48504
/note="22 copies 4 mer acag 72% conserved"
48461. .48504
/note="22 copies 2 mer ac 90% conserved"
49354. .49574
/note="MIR repeat: matches 5. .230 of consensus"
50009. .50539
/note="LIM4 repeat: matches 3878. .4424 of consensus"
51611. .51907
/note="AluX repeat: matches 1. .300 of consensus"
52619. .52658
/note="20 copies 2 mer ca 90% conserved"
52970. .53161

```

Query Match 80.0%; Score 18.4; DB 9; Length 119420;
 Best Local Similarity 95.0%; Pred. No. 3.7e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGAACATACATAACCA 21
 DB 7849 AATGAACATACATAACCA 7830

RESULT 10
 AC073703
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

AC073703 207166 bp DNA linear HTG 18-JUL-2000
 Mus musculus clone RP23-16E5, WORKING DRAFT SEQUENCE, 22 ordered
 pieces.
 AC073703
 AC073703.2 GI:9256762
 HTG; HTGS_PHASE2; HTGS_DRAFT.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 207166)
 DOE Joint Genome Institute.
 Sequencing of Mouse
 Unpublished
 2 (bases 1 to 207166)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Jul 18, 2000 this sequence version replaced gi:8810320.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 1743117
 Center clone name: RCI-23_16E5

Summary Statistics
 Consensus quality: 193525 bases at least Q40
 Consensus quality: 202548 bases at least Q30
 Consensus quality: 204188 bases at least Q20
 Estimated insert size: 208000; agarose-fp estimation
 Estimated insert size: 206166; sum-of-contigs estimation
 Quality coverage: 8.62 in Q20 bases; agarose-fp estimation
 Quality coverage: 8.7 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 4743: contig of 4743 bp in length
 * 4744 4843: gap of unknown length
 * 4844 13587: contig of 8744 bp in length
 * 13588 13687: gap of unknown length
 * 13688 26532: contig of 12845 bp in length
 * 26533 26632: gap of unknown length
 * 26633 39968: contig of 13336 bp in length
 * 39969 40068: gap of unknown length
 * 40069 60997: contig of 20929 bp in length
 * 60998 61097: gap of unknown length
 * 61098 71379: contig of 10282 bp in length
 * 71380 71479: gap of unknown length
 * 71480 93270: contig of 21791 bp in length
 * 93271 93370: gap of unknown length
 * 93371 99451: contig of 6081 bp in length
 * 99452 109513: contig of 9962 bp in length
 * 109514 109613: gap of unknown length
 * 109614 116641: contig of 7028 bp in length
 * 116642 116741: gap of unknown length
 * 116742 128540: contig of 11799 bp in length
 * 128541 128641: gap of unknown length
 * 128641 135192: contig of 6552 bp in length
 * 135193 135292: gap of unknown length

```

* 135293 137898: contig of 2606 bp in length
* 137899 137998: gap of unknown length
* 143409 143408: contig of 5410 bp in length
* 143409 143508: gap of unknown length
* 143509 145880: contig of 2372 bp in length
* 145881 145980: gap of unknown length
* 145981 161155: contig of 15175 bp in length
* 161156 161256: gap of unknown length
* 161257 165123: contig of 3868 bp in length
* 165124 165223: gap of unknown length
* 165224 167221: contig of 1998 bp in length
* 167222 167321: gap of unknown length
* 167322 191494: contig of 24173 bp in length
* 191495 191594: gap of unknown length
* 191595 192681: contig of 1087 bp in length
* 192682 192781: gap of unknown length
* 192782 203772: contig of 10991 bp in length
* 203773 203872: gap of unknown length
* 203873 207166: contig of 3294 bp in length.
FEATURES
    Location/Qualifiers
        1..207166
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone_lib="RP23-16E5"
ORIGIN
    Query Match      80.0%; Score 18.4; DB 2; Length 207166;
    Best Local Similarity 95.0%; Pred. No. 3.5e+02;
    Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AATGAACCTACATACACCA 21
    |||||
Db 93079 AATGACCTACATACACCA 93098

RESULT 11
AC132190/c
LOCUS
DEFINITION Mus musculus clone RP23-378B3, WORKING DRAFT SEQUENCE, 41 unordered
pieces.
AC132190
VERSION AC132190.2 GI:30017734
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nussbaum,C. and Lander,E.
1 (bases 1 to 316678)
Mus musculus, clone RP23-378B3
Unpublished
2 (bases 1 to 316678)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 316678)
Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Liu,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 17, 2003 this sequence version replaced gi:2549775.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----Project Information
Center project name: L26973
Center clone name: 378_B_3
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 906: contig of 906 bp in length
* 907 1006: gap of 100 bp
* 1007 1622: contig of 616 bp in length
* 1623 1722: gap of 100 bp
* 1723 2454: contig of 732 bp in length
* 2455 2554: gap of 100 bp
* 2555 3462: contig of 908 bp in length
* 3463 3562: gap of 100 bp
* 3563 4941: contig of 1378 bp in length
* 4941 5040: gap of 100 bp
* 5041 6356: contig of 1316 bp in length
* 6357 6456: gap of 100 bp
* 6457 7539: contig of 1083 bp in length
* 7540 7639: gap of 100 bp
* 7640 8951: contig of 1312 bp in length
* 8952 9051: gap of 100 bp
* 9052 10202: contig of 1151 bp in length
* 10203 10302: gap of 100 bp
* 10303 11848: contig of 1546 bp in length
* 11849 11948: gap of 100 bp
* 11949 13055: contig of 1107 bp in length
* 13056 13156: gap of 100 bp
* 13156 15527: contig of 2372 bp in length
* 15528 15627: gap of 100 bp
* 15628 17409: contig of 1782 bp in length
* 17410 17509: gap of 100 bp

```

TITLE

```
* 17510 18310: contig of 801 bp in length
* 18311 18410: gap of 100 bp
* 18411 20473: contig of 2063 bp in length
* 20474 20573: gap of 100 bp
* 20574 22133: contig of 1560 bp in length
* 22134 22233: gap of 100 bp
* 22234 24255: contig of 2022 bp in length
* 24256 24356: gap of 100 bp
* 24357 25980: contig of 1625 bp in length
* 25981 26081: gap of 100 bp
* 26082 27721: contig of 1641 bp in length
* 27722 27821: gap of 100 bp
* 27822 30199: contig of 2378 bp in length
* 30200 30299: gap of 100 bp
* 30300 32156: contig of 1856 bp in length
* 32157 34250: gap of 100 bp
* 34251 34350: gap of 100 bp
* 34351 36191: contig of 1841 bp in length
* 36192 36291: gap of 100 bp
* 36292 39231: contig of 2940 bp in length
* 39232 39331: gap of 100 bp
* 39332 43671: contig of 4340 bp in length
* 43672 43771: gap of 100 bp
* 43772 45876: contig of 2105 bp in length
* 45877 45976: gap of 100 bp
* 45977 48712: contig of 2736 bp in length
* 48713 48812: gap of 100 bp
* 48813 52480: contig of 3668 bp in length
* 52481 52580: gap of 100 bp
* 52581 56188: contig of 3608 bp in length
* 56189 56288: gap of 100 bp
* 56289 59456: contig of 3168 bp in length
* 59457 63722: contig of 4166 bp in length
* 63723 63822: gap of 100 bp
* 63823 67556: contig of 3734 bp in length
* 67557 67656: gap of 100 bp
* 67657 72152: contig of 4496 bp in length
* 72153 72252: gap of 100 bp
* 72253 74581: contig of 2329 bp in length
* 74582 74681: gap of 100 bp
* 74682 79961: contig of 5280 bp in length
* 79962 80062: gap of 100 bp
* 80063 85313: contig of 5851 bp in length
* 85314 96395: contig of 10383 bp in length
* 96396 96495: gap of 100 bp
* 96496 142316: contig of 45821 bp in length
* 142317 142416: gap of 100 bp
* 142417 201149: contig of 58733 bp in length
* 201150 201249: gap of 100 bp
* 201250 263398: contig of 62149 bp in length
* 263399 316678: gap of 100 bp
* 316679 53180: contig of 53180 bp in length.
FEATURES
    Location/Qualifiers
    1..316678
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="RP23-378B3"
        /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature 1..906
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left
misc_feature 1007..1622
    /note="assembly_fragment"
misc_feature 1723..2454
    /note="assembly_fragment"
misc_feature 2555..3462
    /note="assembly_fragment"
misc_feature 3563..4940
```

```
misc_feature /note="assembly_fragment"
5041..6356
misc_feature /note="assembly_fragment"
6457..7539
misc_feature /note="assembly_fragment"
7640..8951
misc_feature /note="assembly_fragment"
9052..10202
misc_feature /note="assembly_fragment"
10303..11848
misc_feature /note="assembly_fragment"
11949..13055
misc_feature /note="assembly_fragment"
13156..15527
misc_feature /note="assembly_fragment"
15628..17409
misc_feature /note="assembly_fragment"
17510..18310
misc_feature /note="assembly_fragment"
18411..20473
```

Query Match 80.0%; Score 18.4; DB 2; Length 316678;

Best Local Similarity 95.0%; Pred. No. 3.4e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATGAACTACATCAACCA 21

Db 223361 ATGAACTACATCAACCA 223342

RESULT 12

AC115285/c AC115285 124786 bp DNA linear PRI 29-JUN-2002
LOCUS Homo sapiens chromosome 1 clone RP4-580O19, complete sequence.
DEFINITION AC115285 AL359872
ACCESSION AC115285.2 GI:21629400
VERSION HTG.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 124786)

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,

Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.

and Haugen, E.D.

Direct Submission

Unpublished

REFERENCE

2 (bases 1 to 124786)

AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.

Direct Submission

Submitted (16-MAR-2002) Genome Center, University of Washington,

Box 352145, Seattle, WA 98195, USA

REFERENCE

3 (bases 1 to 124786)

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,

Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.

and Haugen, E.D.

Direct Submission

Submitted (29-JUN-2002) Genome Center, University of Washington,

Box 352145, Seattle, WA 98195, USA

On Jun 29, 2002 this sequence version replaced gi:19526124.

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu>

Contact: uwgchtgs@u.washington.edu

Drafting Center: SC

Center project name: chr-1

Center clone name: RP4-580O19 (sc0808)

Sequencing vector: plasmid; 41% of reads

Sequencing vector: plasmid; 108752; 59% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 124541 bases at least Q40

Consensus quality: 124757 bases at least Q30

Consensus quality: 124780 bases at least Q20

Insert size: 124786; sum-of-contigs

Quality coverage: 6.2x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5: RP1-143P10 AL138837, 17366-bp overlap

3: RP5-947L8 AL355178, 46965-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:

all regions were either double-stranded or sequenced with an

alternate chemistry or covered by high quality data (i.e., Phred

quality > 30); an attempt was made to resolve all sequencing

problems, such as compressions and repeats; all regions were

covered by at least one plasmid subclone or more than one M13

subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest

fingerprinting. Comparison of the experimentally derived digest

fragments with sequence-predicted fragments is given below.

The electronically-digested sequence consists of both insert and

vector, in order to accurately represent the entire circular BAC.

Small fragments below a variable cutoff (approximately 400-800 bp)

are not resolved in the fingerprint and hence do not appear

in the table. There are no significant remaining discrepancies

between the experimental and predicted values. Uniquely ordered

fragments are separated by dashed lines.

NsII				HindIII			
SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt
8243	8134	13702	14292	3464	3373		
3357	3404	5671	5558	449	<800		
266	<800	4488	4326	512	<800		
1629	1613	10240	9842	2814	2849		
9716	9788	2889	3110	1247	1222		
5663	5484	2673	2659	4695	4771		
6253	6313	597	<800	3373	3373		
5429	5484	1973	1984	4630	4595		
10024	9788	3382	3602	1454	1438		
5452	5484	2476	2449	9110	8939		
3066	3090	154	<800	1751	1762		
354	<800	184	<800	5564	5531		
5599	5484	765	757	546	<800		
5566	5484	3932	3853	4849	4771		

FEATURES

source
Location/Qualifiers
1. .124786
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-580019"
/clone_lib="RP4 human PAC library 4"
1. .4
/note="Single subclone region"
40666. .40934
/note="Single subclone region"

misc_feature

misc_feature

1982	1974	3106	3110	2120	2135
8033	8134	4371	4326	1918	1878
7218	7268	2677	2884	2732	2849
138	<800	240	<800	4082	4048
291	<800	4781	4636	4677	4595
101	<800	1080	1085	20100	20083
1703	1688	750	757	3400	3373
1772	1781	868	867	1976	1980
1790	1781	1254	1219	1541	1525
5969	6025	4922	4809	5883	5845
915	924	5624	5558	1458	1438
4017	4098	2416	2449	4218	4243
5400	5484	1156	1085	2799	2849
1871	1856	3663	3853	1946	1980
2265	2311	4011	4074	1862	1878
3440	3404	4336	4326	7423	7360
1633	1613	4122	4326	4309	4243
12054	12157	2445	2449	1129	1139
4103	4098	1993	1984	490	<800
5484	5484	5451	5368	84	<800
		2063	1984	243	<800
		69	<800	84	<800
		301	<800	1405	1438
		3168	3357	10008	10237
		1312	1289	2673	2672
		119	<800	1448	1438
		1772	1726	174	<800
		9736	9395	259	<800
		8431	8183	4752	4771
		1433	1370	1145	1139

genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at [URL: http://www.illumina.com/technology/sequencing/next_generation/sequencing_quality_metrics](http://www.illumina.com/technology/sequencing/next_generation/sequencing_quality_metrics)

FEATURES	source	Location/Qualifiers
misc_feature	1..157653	complement(1..2001 of clone AC004805"
repeat_region	74..284	/rpt_family="MER82"
repeat_region	3408..3501	/rpt_family="L2"
repeat_region	3602..3653	/rpt_family="MIR"
repeat_region	3698..3847	/rpt_family="MIR"
repeat_region	3872..4164	/rpt_family="AluY"
repeat_region	4225..4754	/rpt_family="MLT1E2"
repeat_region	4784..5062	/rpt_family="MLT1E2-internal"
repeat_region	5123..5286	/rpt_family="MLT1E2-internal"
repeat_region	5674..6777	/rpt_family="L1M4C"
repeat_region	6888..8058	/rpt_family="MER113"
repeat_region	8371..8497	/rpt_family="AT-rich"
repeat_region	10516..11013	/rpt_family="MLT1E2"
repeat_region	14787..14807	/rpt_family="AT-rich"
STS	16143..16381	/standard_name="D12S340"
STS	16196..16315	/standard_name="D12S340"
repeat_region	16260..16292	/rpt_family="(TG)n"
STS	17707..17926	/standard_name="A007A34"
STS	17815..17945	/standard_name="SGC31838"
repeat_region	18920..19201	/rpt_family="AluSq"
STS	21082..21265	/standard_name="RH93075"
repeat_region	22757..23045	/rpt_family="AluSx"

```

repeat_region 25096..25120
                /rpt_family="AT_rich"

Query Match      79.1%; Score 18.2; DB 9; Length 157653;
Best Local Similarity 87.0%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GAATGAACATACATAACACACC 23
        ||||| ||||| ||||| |||||
Db      3457 GAATGAACATACATAACACACC 3479

RESULT 14
AC025303
LOCUS
DEFINITION
Homo sapiens clone RP11-689C19, WORKING DRAFT SEQUENCE, 33
unordered pieces.
AC025303
AC025303.1 GI:7209939
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Birten,B., Linton,L., Nusbaum,C. and Lander,E.
  (bases 1 to 202398)
  Homo sapiens, clone RP11-689C19
  Unpublished
  2 (bases 1 to 202398)
  Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
  Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
  Boguslavskiy,L., Boukigalter,B., Brown,A., Burkett,G.,
  Campopiano,A., Castle,A., Choepel,J., Colangelo,M., Collins,S.,
  Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
  Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
  Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
  Grant-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
  Howland,J.R.C., Iliev,I., Johnson,R., Jones,C., Kann,D., Karatas,A.,
  Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
  Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
  McCarthy,M., McEwan,P., McGurk,A., McKernan,K., MCPheeters,R.,
  Meldrim,J., Meneses,L., Mihova,I., Miranda,C., Mienga,V., Morrow,J.,
  Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
  O'Neill,D., Olliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
  Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
  Roy,A., Santos,R., Schajnovic,N., Subramanian,A., Talamas,J.,
  Stange-Thomann,N., Stejnar,J., Tirrell,A., Travers,M., Trigilio,J.,
  Tsefaye,S., Theodore,J., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.,J.,
  Vassiliev,H., Viel,R., Vo,A., Zimmer,A. and Zody,W.
  Young,G., Zainoun,J.,
  Direct Submission
  Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L4340
  Center clone name: 689 C19
  ----- Summary Statistics
  Sequencing vector: M13; M77815; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.960731
  Consensus quality: 174122 bases at least Q40
  Consensus quality: 186313 bases at least Q30
  Consensus quality: 192877 bases at least Q20
  Insert size: 194000; agarose-fp
  Insert size: 199198; run-of-contigs

```


Quality coverage: 3.5 in Q20 bases; agarose-fp
 Quality coverage: 3.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 33 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1395: contig of 1395 bp in length
 1396: gap of 100 bp
 1496: contig of 1257 bp in length
 2752: gap of 100 bp
 2852: contig of 1570 bp in length
 4422: gap of 100 bp
 4522: contig of 2238 bp in length
 6760: gap of 100 bp
 6860: gap of 100 bp
 6861: contig of 1849 bp in length
 8709: gap of 100 bp
 8809: contig of 2436 bp in length
 11245: gap of 100 bp
 11345: contig of 2103 bp in length
 13448: gap of 100 bp
 13449: gap of 100 bp
 13449: contig of 2424 bp in length
 15972: gap of 100 bp
 16072: contig of 3020 bp in length
 19092: gap of 100 bp
 19192: contig of 2853 bp in length
 22045: gap of 100 bp
 22145: gap of 100 bp
 25821: contig of 3676 bp in length
 25822: gap of 100 bp
 30209: contig of 4288 bp in length
 30309: gap of 100 bp
 33625: contig of 3316 bp in length
 33725: gap of 100 bp
 38386: contig of 4661 bp in length
 38486: gap of 100 bp
 43134: contig of 4648 bp in length
 43234: gap of 100 bp
 47752: contig of 4518 bp in length
 47852: gap of 100 bp
 52841: contig of 4989 bp in length
 52941: gap of 100 bp
 58193: contig of 5252 bp in length
 58293: gap of 100 bp
 63614: contig of 5321 bp in length
 63714: gap of 100 bp
 69041: contig of 5327 bp in length
 69141: gap of 100 bp
 75114: contig of 5973 bp in length
 75214: gap of 100 bp
 82031: contig of 6817 bp in length
 82131: gap of 100 bp
 87882: contig of 5751 bp in length
 87982: gap of 100 bp
 95376: contig of 7394 bp in length
 95476: gap of 100 bp
 102941: contig of 7465 bp in length
 103041: gap of 100 bp
 111576: contig of 8535 bp in length
 111577: gap of 100 bp
 120340: contig of 8664 bp in length
 120441: gap of 100 bp
 131444: contig of 11004 bp in length
 131445: gap of 100 bp
 143154: contig of 11646 bp in length
 143190: gap of 100 bp
 143290: contig of 13055 bp in length
 156345: gap of 100 bp
 156445: contig of 13463 bp in length
 169908: gap of 100 bp
 170008: contig of 13463 bp in length
 170009: gap of 100 bp

* 170009 186176: contig of 16168 bp in length
 * 186177 186276: gap of 100 bp
 * 186277 202398: contig of 16122 bp in length.
 FEATURES
 source
 1. 202398
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-689C19"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1395
 /note="assembly_fragment"
 1496. 2752
 /note="assembly_fragment"
 2853. 4422
 /note="assembly_fragment"
 4523. 6760
 /note="assembly_fragment"
 6861. 8709
 /note="assembly_fragment"
 clone end: T7
 vector side: left
 8810. 11245
 /note="assembly_fragment"
 11345. 13448
 /note="assembly_fragment"
 13549. 15972
 /note="assembly_fragment"
 16073. 19092
 /note="assembly_fragment"
 19193. 22045
 /note="assembly_fragment"
 22146. 25821
 /note="assembly_fragment"
 25922. 30209
 /note="assembly_fragment"
 30310. 33625
 /note="assembly_fragment"
 33726. 38386
 /note="assembly_fragment"
 38487. 43134
 /note="assembly_fragment"
 43235. 47752
 /note="assembly_fragment"
 47853. 52841
 /note="assembly_fragment"
 52942. 58193
 /note="assembly_fragment"
 58294. 63614
 /note="assembly_fragment"
 63715. 69041
 /note="assembly_fragment"
 69142. 75114
 /note="assembly_fragment"
 75215. 82031
 /note="assembly_fragment"
 82132. 87882
 /note="assembly_fragment"
 87983. 95376
 /note="assembly_fragment"
 95477. 102941
 /note="assembly_fragment"
 103042. 111576
 /note="assembly_fragment"
 111677. 120340
 /note="assembly_fragment"
 120441. 131444
 /note="assembly_fragment"

Query Match 79.1%; Score 18.2; DB 2; Length 202398;
 Best Local Similarity 87.0%; Pred. No. 4.3e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

[illegible]

```

* * 72154 75289: gap of unknown length
* * contig of 3136 bp in length
* * 75290 78389: gap of unknown length
* * contig of 3100 bp in length
* * 78390 80910: gap of unknown length
* * contig of 2521 bp in length
* * 80911 82153: gap of unknown length
* * contig of 1243 bp in length
* * 82154 84007: gap of unknown length
* * contig of 1854 bp in length
* * 84008 86443: gap of unknown length
* * contig of 2436 bp in length
* * 86444 89640: gap of unknown length
* * contig of 3197 bp in length
* * 89641 91024: gap of unknown length
* * contig of 1384 bp in length
* * 91025 93347: gap of unknown length
* * contig of 2323 bp in length
* * 93348 95487: gap of unknown length
* * contig of 2140 bp in length
* * 95488 98489: gap of unknown length
* * contig of 3002 bp in length
* * 98490 100499: gap of unknown length
* * contig of 2010 bp in length
* * 100500 104757: gap of unknown length
* * contig of 4258 bp in length
* * 104758 106505: gap of unknown length
* * contig of 1748 bp in length
* * 106506 112258: gap of unknown length
* * contig of 5753 bp in length
* * 112259 115111: gap of unknown length
* * contig of 2853 bp in length
* * 115112 117906: gap of unknown length
* * contig of 2795 bp in length
* * 117907 120203: gap of unknown length
* * contig of 2297 bp in length
* * 120204 125155: gap of unknown length
* * contig of 4952 bp in length
* * 125156 132035: gap of unknown length
* * contig of 6880 bp in length
* * 132036 138866: gap of unknown length
* * contig of 6831 bp in length
* * 138867 146355: gap of unknown length
* * contig of 7489 bp in length
* * 146356 155442: gap of unknown length
* * contig of 9087 bp in length
* * 155443 161962: gap of unknown length
* * contig of 6520 bp in length
* * 161963 170957: gap of unknown length
* * contig of 8995 bp in length
* * 170958 181490: gap of unknown length
* * contig of 10533 bp in length
* * 181491 193728: gap of unknown length
* * contig of 12238 bp in length
* * 193729 204122: gap of unknown length
* * contig of 10394 bp in length
* * 204123 223034: gap of unknown length
* * contig of 18912 bp in length
* * 223035 238554: gap of unknown length
* * contig of 15520 bp in length.
FEATURES Location/Qualifiers
Query Match 79.1%; Score 18.2; DB 2; Length 238554;
Best Local Similarity 87.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAATGAACATACATACACC 23
Db 9289 GAATGAACATACATACACC 9311

```

Search completed: March 24, 2004, 23:01:21
 Job time : 515.584 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 18:56:32 ; Search time 120.427 Seconds
(without alignments)
811.351 Million cell updates/sec

Title: US-09-889-611A-19

Perfect score: 23

Sequence: 1 gaatgaactacatacaaccacc 23

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	23	100.0	23	3	AAA71452 Human meg
2	23	100.0	1431	3	AAA71435 Human meg
3	22	95.7	30	3	AAA71448 Human meg
4	22	95.7	128	3	AAA71434 Human meg
5	20.4	88.7	6182	6	ABL34015 Human imm
6	17.2	74.8	350	5	AAS34210 Human cdn
7	17.2	74.8	580	6	ABQ23445 Oligonuc
8	17.2	74.8	1350	5	AAS72599 DNA encod
9	17.2	74.8	4789	9	ADD47985 Rat gene
10	17.2	74.8	9087	6	ABK31277 Signal tr
11	17.2	74.8	9087	6	ABL70238 Chemical
12	17.2	74.8	9087	6	AAS61181 Human gen
13	17.2	74.8	12879	6	ABK92230 Prostata
14	17.2	74.8	13202	4	AAK51828 Human pol
15	17.2	74.8	16983	9	ADD47173 Rat gene
16	17.2	74.8	16983	9	ADD47173 Rat gene
17	17	73.9	16918	6	ABL33617 Human imm
18	16.8	73.0	744	7	ACF69209 Photorhab
19	16.8	73.0	1215	7	ACA39407 Prokaryot
20	16.8	73.0	2477	4	ABL27492 Drosophil
21	16.8	73.0	2815	4	AAK91396 Human col
22	16.8	73.0	2815	4	AAI57764 Human col
23	16.8	73.0	2815	6	ABS99941 Genomic D

24	16.8	73.0	2815	9	ADB93094	ADB93094 Human col
25	16.8	73.0	2851	4	AAK91397	AAK91397 Human dig
26	16.8	73.0	2851	4	AAK91395	AAK91395 Human dig
27	16.8	73.0	2851	4	AAI57763	AAI57763 Human col
28	16.8	73.0	2851	4	AAI57765	AAI57765 Human col
29	16.8	73.0	2851	6	ABS99940	ABS99940 Genomic D
30	16.8	73.0	2851	6	ABS99942	ABS99942 Genomic D
31	16.8	73.0	2851	9	ADB93093	ADB93093 Human col
32	16.8	73.0	2851	9	ADB93095	ADB93095 Human col
33	16.8	73.0	8262	4	AAK71151	AAK71151 Human imm
34	16.8	73.0	9965	6	ABL33526	ABL33526 Human imm
35	16.8	73.0	110000	2	AAT58840_1	Continuation (2 of
36	16.8	73.0	110000	7	ACF67367_21	Continuation (22 o
37	16.8	73.0	243072	7	ACF65382	ACF65382 Photorhab
38	16.6	72.2	219	3	AAC22299	AAC22299 Human sec
39	16.6	72.2	390	7	ACA48438	ACA48438 Prokaryot
40	16.6	72.2	461	7	ABZ55960	ABZ55960 Aspergill
41	16.6	72.2	664	4	AAI20886	AAI20886 Human bre
42	16.6	72.2	4569	4	ABL15988	ABL15988 Drosophil
43	16.6	72.2	17993	6	AAK46721	AAK46721 Human tra
44	16.6	72.2	24978	2	AAK60209	AAK60209 SEQ ID 3
45	16.6	72.2	24979	3	AAA52321	AAA52321 Genomic D

ALIGNMENTS

RESULT 1
AAA71452
ID AAA71452 standard; DNA; 23 BP.
XX
AC AAA71452;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter PCR primer SEQ ID NO: 19.
XX
KW Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megsin useful for screening proteins.
XX
PS Example 4; Page 42; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors).
XX
CC AAA71434-A71469 represent PCR primers used in the method described in the invention
XX
SQ Sequence 23 BP; 11 A; 7 C; 2 G; 3 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GAATGAACACTACATAACACACC 23
DB 1 GAATGAACACTACATAACACACC 23

RESULT 2
AAA71435
AC AAA71435;
XX
XX 01-DEC-2000 (first entry)
XX
XX Human megsin promoter fragment DNA.
XX
XX Promoter; megsin; human; protein isolation; screening. ss.
XX
XX Homo sapiens.
XX
XX WO2000043528-A1.
XX
XX 27-JUL-2000.
XX
XX 25-JAN-2000; 2000WO-JP000350.
XX
XX 25-JAN-1999; 99JP-00015667.
XX
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX
XX Miyata T;
XX
XX WPI; 2000-543257/49.
XX
XX DNA for promoter region of megsin useful for screening proteins.
XX
XX Example 5; Page 40; 45pp; Japanese.
XX
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the
XX above vector; and (3) protein produced using (I). (I) is useful for
XX screening and isolating proteins (especially transcription factors).
XX
XX AAA71434-A71469 represent PCR primers used in the method described in the
XX invention
XX
XX Sequence 30 BP; 13 A; 8 C; 3 G; 6 T; 0 U; 0 Other;
XX
XX Query Match 95.7%; Score 22; DB 3; Length 30;
XX Best Local Similarity 100.0%; Pred. No. 3;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AATGAACACTACATAACACACC 23
DB 1 AATGAACACTACATAACACACC 22

RESULT 4
AAA71434
ID AAA71434 standard; DNA; 128 BP.
XX
XX AC AAA71434;
XX
XX DT 01-DEC-2000 (first entry)
XX
XX DE Human megsin promoter fragment DNA.
XX
XX KW Promoter; megsin; human; protein isolation; screening. ss.
XX
XX OS Homo sapiens.
XX
XX WO2000043528-A1.
XX
XX PD 27-JUL-2000.
XX
XX PF 25-JAN-2000; 2000WO-JP000350.
XX
XX PR 25-JAN-1999; 99JP-00015667.
XX
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX
XX PI Miyata T;
XX
XX WPI; 2000-543257/49.
XX
XX DNA for promoter region of megsin useful for screening proteins.
XX
XX Claim 1; Page 32; 45pp; Japanese.
XX
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the

```

CC above vector; and (3) protein produced using (1). (1) is useful for
 CC screening and isolating proteins (especially transcription factors). This
 CC sequence represents the human meglin promoter which is described in the
 CC method of the invention

XX Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
 SQ Query Match 95.7%; Score 22; DB 3; Length 128;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAACACTACATAACACACC 23
 DB 1 AATGAACACTACATAACACACC 22

RESULT 5
 ID ABL34015/C
 AC ABL34015;
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 1988.
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anaemic; cytostatic; neutropenic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.

XX Homo sapiens.
 XX WO200200928-A2.
 XX 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP007537.
 XX 30-JUN-2000; 2000DE-01032529.
 XX 01-SEP-2000; 2000DE-01043826.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

PS Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;
 SQ Query Match 88.7%; Score 20.4; DB 6; Length 6182;
 Best Local Similarity 95.5%; Pred. No. 22;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGAACACTACATAACACACC 23
 DB 1310 AATAAACTACATAACACACC 1289

RESULT 6
 AAS34210/C
 ID AAS34210 standard; cDNA; 350 BP.
 XX AAS34210;
 AC AAS34210;
 DT 17-DEC-2001 (first entry)
 DE Human cDNA encoding a novel foetal antigen, SEQ ID No 734.

XX Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
 KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
 KW cardiovascular; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; breast neoplasm; cancer;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; food additive.

XX Homo sapiens.
 OS WO200155312-A2.
 PN 02-AUG-2001.
 XX 17-JAN-2001; 2001WO-US001321.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246529P.
 PR 08-NOV-2000; 2000US-0246603P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 05-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254037P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-488782/53.
 DR P-PSDB; AAU21390.
 DR
 XX
 PT New polynucleotides and polypeptides for diagnosing, treating, preventing
 PT or prognosing e.g. diseases or disorders of the nervous, musculoskeletal,
 PT excretory, gastrointestinal, reproductive, and respiratory systems.
 XX
 PS Claim 1; SEQ ID NO 734; 642pp; English.
 XX
 CC The invention relates to novel nucleic acids encoding novel human foetal
 CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
 CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. The antibodies to the antigens can also be used
 CC in alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities, fat content, lipid, protein, carbohydrate, Numerous
 CC vitamins, minerals, cofactors and other nutritional components. Numerous
 CC examples of diseases and disorders treated by the nucleic acids and
 CC proteins are given in the specification. The present sequence encodes a
 CC foetal antigen of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in

Query Match 74.8%; Score 17.2; DB 5; Length 350;
 Best Local Similarity 86.4%; Pred. No. 4.6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AATGAACCTACATACACCACC 23

Db 145 AATGAACCTCATTACATCACC 124

RESULT 7
 ABQ23445

```

ID  ABQ23445 standard; DNA; 580 BP.
XX  ABQ23445;
XX  12-JUL-2002 (first entry)
XX  Oligonucleotide for detecting cytosine methylation SEQ ID NO 10036.
XX  Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX  drug; side effect; cancer; central nervous system; cardiovascular;
XX  gastrointestinal; respiratory system; single nucleotide polymorphism;
XX  SNP; cell differentiation; ds.
XX  Homo sapiens.
XX  WO200218632-A2.
XX  07-MAR-2002.
XX  01-SEP-2001; 2001WO-EP010074.
XX  01-SEP-2000; 2000DE-01043826.
XX  05-SEP-2000; 2000DE-01044543.
XX  (EPIG-) EPIGENOMICS AG.
XX  Olek A, Piepenbrock C, Berlin K, Guetig D;
XX  WPI; 2002-371829/40.
XX  Determining the degree of cytosine methylation in genomic DNA, useful for
XX  diagnosis and prognosis, comprises selective hybridization of amplicons
XX  from chemically treated DNA.
XX  Claim 12; 56pp + Sequence Listing; 56pp; German.
XX  This invention describes a novel method for determining the degree of
XX  methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX  genomic sample of DNA. The sample is treated chemically to convert
XX  cytosine (C) but not methylated C, to uracil, then part of the genomic
XX  DNA that contains the target C is amplified to form a labeled amplicon.
XX  The amplicon is hybridised to two classes, each with at least one member,
XX  of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX  degree of hybridisation to both classes is determined from the label on
XX  the amplicon. From the ratio of labels hybridised to the two classes of
XX  oligomers, the degree of methylation is calculated. The method is used:
XX  (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX  and of a wide range of diseases, e.g. cancer, disorders of the central
XX  nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX  particularly by detecting mutations or single nucleotide polymorphisms
XX  (SNP's); and (ii) for differentiation of cell or tissue types and for
XX  investigating cell differentiation. The method allows the methylation
XX  status of many C residues to be determined simultaneously. ABQ13410-
XX  ABQ54121 represent genomic DNA sequences used to illustrate the method
XX  for determining the degree of cytosine methylation described in the
XX  disclosure of the invention
XX  Sequence 580 BP; 256 A; 150 C; 49 G; 125 T; 0 U; 0 Other;
XX  Query Match 74.8%; Score 17.2; DB 6; Length 580;
XX  Best Local Similarity 86.4%; Pred. No. 4.8e+02;
XX  Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX  QY 2 AATGAACACTACATCAACGCC 23
XX  DB 254 AATCAACTACATCAACGCC 275
XX  RESULT 8
XX  ABQ23444/c
XX  ID ABQ23444 standard; DNA; 580 BP.
XX  AC ABQ23444;

```

```

XX  12-JUL-2002 (first entry)
XX  Oligonucleotide for detecting cytosine methylation SEQ ID NO 10035.
XX  Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX  drug; side effect; cancer; central nervous system; cardiovascular;
XX  gastrointestinal; respiratory system; single nucleotide polymorphism;
XX  SNP; cell differentiation; ds.
XX  Homo sapiens.
XX  WO200218632-A2.
XX  07-MAR-2002.
XX  01-SEP-2001; 2001WO-EP010074.
XX  01-SEP-2000; 2000DE-01043826.
XX  05-SEP-2000; 2000DE-01044543.
XX  (EPIG-) EPIGENOMICS AG.
XX  Olek A, Piepenbrock C, Berlin K, Guetig D;
XX  WPI; 2002-371829/40.
XX  Determining the degree of cytosine methylation in genomic DNA, useful for
XX  diagnosis and prognosis, comprises selective hybridization of amplicons
XX  from chemically treated DNA.
XX  Claim 12; 56pp + Sequence Listing; 56pp; German.
XX  This invention describes a novel method for determining the degree of
XX  methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX  genomic sample of DNA. The sample is treated chemically to convert
XX  cytosine (C) but not methylated C, to uracil, then part of the genomic
XX  DNA that contains the target C is amplified to form a labeled amplicon.
XX  The amplicon is hybridised to two classes, each with at least one member,
XX  of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX  degree of hybridisation to both classes is determined from the label on
XX  the amplicon. From the ratio of labels hybridised to the two classes of
XX  oligomers, the degree of methylation is calculated. The method is used:
XX  (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX  and of a wide range of diseases, e.g. cancer, disorders of the central
XX  nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX  particularly by detecting mutations or single nucleotide polymorphisms
XX  (SNP's); and (ii) for differentiation of cell or tissue types and for
XX  investigating cell differentiation. The method allows the methylation
XX  status of many C residues to be determined simultaneously. ABQ13410-
XX  ABQ54121 represent genomic DNA sequences used to illustrate the method
XX  for determining the degree of cytosine methylation described in the
XX  disclosure of the invention
XX  Sequence 580 BP; 125 A; 49 C; 150 G; 256 T; 0 U; 0 Other;
XX  Query Match 74.8%; Score 17.2; DB 6; Length 580;
XX  Best Local Similarity 86.4%; Pred. No. 4.8e+02;
XX  Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX  QY 2 AATGAACACTACATCAACGCC 23
XX  DB 327 AATCAACTACATCAACGCC 306
XX  RESULT 9
XX  AAS72599/c
XX  ID AAS72599 standard; cDNA; 1350 BP.
XX  AC AAS72599;
XX  DT 13-FEB-2002 (first entry)
XX

```


DE DNA encoding novel human diagnostic protein #8403.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-FSDB; ABG08412.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 8403; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1350 BP; 374 A; 400 C; 171 G; 404 T; 0 U; 1 Other;
Query Match 74.8%; Score 17.2; DB 5; Length 1350;
Best Local Similarity 86.4%; Pred. No. 5.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AATGAACACTACATACACACCACC 23
|||||
Db 1018 AATGATCTACACAAACACCACC 997
|||||
RESULT 10
ADD47985/c
ID ADD47985 standard; DNA; 4789 BP.
XX
XX AC ADD47985;
XX
XX DT 29-JAN-2004 (first entry)
XX
XX DE Rat gene U30813, SEQ ID NO 13681.
XX

KW Rat; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX
XX Rattus norvegicus.
XX WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX (GEHO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX DR GENBANK; U30813.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4789 BP; 1333 A; 981 C; 1050 G; 1425 T; 0 U; 0 Other;
Query Match 74.8%; Score 17.2; DB 9; Length 4789;
Best Local Similarity 86.4%; Pred. No. 5.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AATGAACACTACATACACACCACC 23
|||||
Db 2746 AATGACACATGCCACACCACC 2725
|||||
RESULT 11
ABK31277/c
ID ABK31277 standard; DNA; 9087 BP.
XX


```
XX OS Homo sapiens.
XX PN WO200177375-A2.
XX PD 18-OCT-2001.
XX PF 06-APR-2001; 2001WO-EP003368.
XX PR 06-APR-2000; 2000DE-01019058.
XX PR 07-APR-2000; 2000DE-01019173.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043626.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-017470/02.
XX PT New nucleic acid sequences from chemically modified genes associated with
XX gene regulation, useful for analyzing cytosine methylations for diagnosis
XX and therapy of diseases e.g. severe combined immunodeficiency disease.
XX PS Claim 1; SEQ ID NO 140; 26pp; English.
XX CC The invention relates to 224 nucleic acid sequences comprising at least
XX 18 bases of a chemically pretreated gene associated with gene regulation
XX selected from 43 known genes (or complementary sequences). The chemical
XX pretreatment converts cytosine bases unmethylated at the 5-position to
XX uracil or another base with hybridisation behaviour dissimilar to
XX cytosine, to enable analysis of cytosine methylations. The DNA sequences,
XX oligomers (or sets/arrays) and method are useful in the diagnosis of
XX diseases (or predisposition to diseases) associated with gene regulation
XX and in therapy of such diseases, by enabling analysis of the cytosine
XX methylation patterns of such genes, kits are provided. They are
XX especially useful in diagnosis and therapy of e.g. severe combined
XX immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
XX and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
XX syndrome, renal disease, preeclampsia, graft versus-host disease. The
XX present sequence is a sequence included in the sequence data for this
XX specification and is associated with the human gene regulation-associated
XX genes. Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 9087 BP; 2207 A; 253 C; 2592 G; 4034 T; 0 U; 1 Other;
Query Match 74.8%; Score 17.2; DB 6; Length 9087;
Best Local Similarity 86.4%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 AATGAACCTACATACCAACACC 23
Db 2025 AACGAACCTACATCAACACC 2004
RESULT 14
ABK92230/c
ID ABK92230 standard; DNA; 12879 BP.
XX AC ABK92230;
XX DT 15-AUG-2002 (first entry)
XX DE Prostate cancer-associated DNA sequence #116.
XX KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
XX gene therapy; gene; ds.
XX OS Mammalia.
XX PN WO200230268-A2.
```

```
XX PD 18-APR-2002.
XX PF 12-OCT-2001; 2001WO-US032045.
XX PR 13-OCT-2000; 2000US-00687576.
XX PR 08-DEC-2000; 2000US-00733288.
XX PR 08-DEC-2000; 2000US-00733742.
XX PR 24-JAN-2001; 2001US-0263957P.
XX PR 16-MAR-2001; 2001US-0276791P.
XX PR 16-MAR-2001; 2001US-0276888P.
XX PR 06-APR-2001; 2001US-0281922P.
XX PR 24-APR-2001; 2001US-0286214P.
XX PR 30-APR-2001; 2001US-00847046.
XX PR 04-MAY-2001; 2001US-0288589P.
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX DR WPI; 2002-471335/50.
XX DR P-PSDB; ABG61913.
XX PT Detecting a prostate cancer-associated transcript in a cell in a patient,
XX useful for diagnosing prostate cancer (PC) or screening modulators of PC,
XX by determining if prostate cancer-associated genes are expressed in a
XX prostate tissue.
XX PS Claim 22; Page 394-397; 436pp; English.
XX CC The present invention relates to methods of detecting a prostate cancer-
XX associated transcript in a cell from a patient. The method comprises
XX contacting a biological sample from the patient with prostate cancer-
XX associated polynucleotides (designated PC genes) that selectively
XX hybridise to a sequence that is at least 80% identical to them. The
XX prostate cancer-associated polynucleotide sequences are differentially
XX expressed in prostate tumour tissue or in prostate cancer and are derived
XX from the tissues of various organisms such as humans or other mammals
XX (e.g. mice, sheep and dogs). The methods of the invention are useful for
XX diagnosing and treating prostate cancer in mammals. The prostate cancer-
XX associated genes are useful for diagnosing or treating prostate cancer,
XX as well as for identifying modulators of prostate cancer or agents that
XX inhibit prostate cancer. The nucleic acid sequences are particularly
XX useful in gene therapy, as a vaccine or in antisense applications.
XX CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
XX sequences
XX SQ Sequence 12879 BP; 3852 A; 2674 C; 2768 G; 3585 T; 0 U; 0 Other;
Query Match 74.8%; Score 17.2; DB 6; Length 12879;
Best Local Similarity 86.4%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 AATGAACCTACATACCAACACC 23
Db 11765 AATGAACCTACATCAACACC 11744
RESULT 15
AAK51828/c
ID AAK51828 standard; cDNA; 13202 BP.
XX AC AAK51828;
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 373.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX PN
```

OS Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US004098.
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX 20-JUN-2000; 2000US-00598075.
XX 19-JUL-2000; 2000US-00620325.
XX 01-SEP-2000; 2000US-00654936.
XX 15-SEP-2000; 2000US-00663561.
XX 20-OCT-2000; 2000US-00693325.
XX 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
XX P-PSDB; AAM78695.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX Claim 1; Page 1414-1426; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication
XX Sequence 13202 BP; 3921 A; 2764 C; 2865 G; 3652 T; 0 U; 0 Other;
XX Query Match 74.8%; Score 17.2; DB 4; Length 13202;
XX Best Local Similarity 86.4%; Pred. No. 5.9e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AATGAACACTACATCAACACCACC 23
DB 12091 AATGAACCTCCATTACATCACC 12070
Search completed: March 24, 2004, 22:28:19
Job time : 124.427 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:57:32 ; Search time 1119.76 Seconds
(without alignments)
613.370 Million cell updates/sec

Title: US-09-889-611A-19

Perfect score: 23
Sequence: 1 gaatgaactacatacaaccacc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estci:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vri:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.4	80.0	569	28	CC162058
C 2	18.4	80.0	684	28	AZ400233
C 3	18.4	80.0	782	29	CG114080
C 4	18.4	80.0	785	28	CC090422

5	18.4	80.0	927	28	CC134385
6	18.4	80.0	999	10	BE572261
C 7	18.4	80.0	1201	9	AL572872
C 8	18.2	79.1	129	10	AW698814
C 9	18.2	79.1	307	14	CD183415
C 10	18.2	79.1	309	14	CD133709
C 11	18.2	79.1	313	14	CD188719
C 12	18.2	79.1	317	14	CD165299
C 13	18.2	79.1	531	14	CD189864
C 14	18.2	79.1	537	14	CD083615
C 15	18.2	79.1	556	14	CD073220
C 16	18.2	79.1	648	14	CD074368
C 17	18.2	79.1	762	14	CF443253
C 18	18.2	79.1	763	10	BE387554
C 19	18.2	79.1	813	29	CG442765
C 20	18.2	79.1	845	29	CG442758
C 21	18.2	79.1	294	10	BF365378
C 22	17.8	77.4	72	14	CF608248
C 23	17.8	77.4	180	28	AZ442459
C 24	17.8	77.4	217	9	AI901912
C 25	17.8	77.4	247	9	AV208603
C 26	17.8	77.4	333	28	AQ908428
C 27	17.8	77.4	431	28	AQ680312
C 28	17.8	77.4	456	13	BY548796
C 29	17.8	77.4	503	28	BH193866
C 30	17.8	77.4	625	28	BZ484996
C 31	17.8	77.4	632	28	BH844430
C 32	17.8	77.4	633	28	BH509041
C 33	17.8	77.4	649	14	CF207448
C 34	17.8	77.4	652	28	AZ606165
C 35	17.8	77.4	666	28	BH247014
C 36	17.8	77.4	682	28	BH841120
C 37	17.8	77.4	694	28	BZ005096
C 38	17.8	77.4	699	28	BH655941
C 39	17.8	77.4	701	28	BH960009
C 40	17.8	77.4	708	28	AZ008960
C 41	17.8	77.4	713	28	BZ678717
C 42	17.8	77.4	766	28	BH90439
C 43	17.8	77.4	782	14	CF066855
C 44	17.8	77.4	820	28	BZ487245
C 45	17.8	77.4	890	28	BZ389277

ALIGNMENTS

RESULT 1
CC162058/c
LOCUS
DEFINITION
CC162058
WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CC162058
ii81b04.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Accession
Version
Keywords
Source
Organism
REFERENCE
AUTHORS

Seq primer: -21M13UnivFwd
 Class: shotgun
 High quality sequence stop: 569.
 Location/Qualifiers

FEATURES

1. 569
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="ii81b04"
 /lab_host="DH5a"
 /clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
 The vector was digested with XbaI and one nucleotide was
 added by fill in the recessive 3' end. The genomic DNA
 was nebulized, end repaired, adaptor ligated and size
 fractionated using sephadex. The resulting fragments were
 between 0.8 and 3 kb and were cloned into the vector (.x/y
 reads in M13mp19, .b/g reads in pUC19). The same ligation
 was transformed into DH5a."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 569;
 Best Local Similarity 95.0%; Pred. No. 2.6e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 4 TGAACACTACATAACCAACCACC 23
 ||||||||||||||||||
 Db 59 TGGACTACATACCAACCACC 40

RESULT 2

AZ400233 684 bp DNA linear GSS 03-OCT-2000
 LOCUS
 DEFINITION 1M0165121F Mouse 10kb plasmid UUC1M library Mus musculus genomic
 clone UUC1M0165121 F, genomic survey sequence.

ACCESSION

VERSION
 AZ400233.1 GI:10515307

KEYWORDS

GSS.

SOURCE

Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 684)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A. and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

COMMENT
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0166 ROW: L COLUMN: 21

Seq primer:

CGTTGTAACACGCGCCAGT

Class:

plasmid ends

High quality sequence stop:

684.

Location/Qualifiers

1. 684
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUC1M0165121"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

FEATURES

source

/clone_lib="Mouse 10kb plasmid UUC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (GI4732114|GB|AF12072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 684;
 Best Local Similarity 95.0%; Pred. No. 2.6e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 AATGAACACTACATAACCAACCA 21
 ||||||||||||||||||
 Db 393 AATGAACACTACATAACCAACCA 412

RESULT 3

CG114080/c 782 bp DNA linear GSS 20-AUG-2003
 LOCUS
 DEFINITION PUJGG81TD ZM_0_6_1_0 KB Zea mays genomic clone ZMBTa0661N17,
 genomic survey sequence.

ACCESSION

VERSION
 CG114080.1 GI:33997517

KEYWORDS

GSS.

SOURCE

Zea mays
 Zea mays
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 782)
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Reenick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

REFERENCE

AUTHORS

TITLE

Maize Genomics Consortium

JOURNAL

COMMENT
 Unpublished (2003)

Other GSSs:

PUJGG81TB

Contact:

Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel:

301-838-5843

Fax:

301-838-0208

Email:

whitelaw@tigr.org

Seq primer:

TF

Class:

sheared ends.

Location/Qualifiers

1. 782
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMBTa0661N17"
 /clone_lib="ZM_0_6_1_0 KB"
 /note="vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN

Query Match 80.0%; Score 18.4; DB 29; Length 782;
 Best Local Similarity 95.0%; Pred. No. 2.6e+03;

```

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TGAAGTACATACACACACC 23
    |||||
Db 500 TGGACTACATACACACACC 481

CC090422 785 bp DNA linear GSS 16-APR-2003
CSU-K33r.9P3.T7 CSU-K33r Aedes aegypti genomic clone CSU-K33r.9P3,
genomic survey sequence.
CC090422
CC090422.1 GI:29947659
GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
Stegomyia.
REFERENCE 1 (bases 1 to 785)
AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE End sequencing of Aedes aegypti BACs
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CSU-K33r.9P3.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado
State University.
Seq primer: T7
Class: BAC ends.

FEATURES
source
LOCATION/Qualifiers
1..785
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/strain="Rexville"
/db_xref="taxon:7159"
/clone="CSU-K33r.9P3"
/clone_lib="CSU-K33r"
/notes="Vector: pBelBAC11; Site_1: HindIII"

ORIGIN
Query Match 80.0%; Score 18.4; DB 28; Length 785;
Best Local Similarity 95.0%; Pred. No. 2.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATGAAGTACATACACACC 20
    |||||
Db 380 GAATGAAGTACATACACACC 361

CC134385 927 bp DNA linear GSS 16-APR-2003
NDL.60D18.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
NDL.60D18, genomic survey sequence.
CC134385
CC134385.1 GI:30003440
GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
Stegomyia.
REFERENCE 1 (bases 1 to 927)
AUTHORS Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE BAC end sequencing of Aedes aegypti
JOURNAL Unpublished (2003)

```

```

Other_GSSs: NDL.60D18.T7
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by David Severson
Seq primer: SP6
Class: BAC ends.

FEATURES
source
LOCATION/Qualifiers
1..927
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/strain="liverpool"
/db_xref="taxon:7159"
/clone="NDL.60D18"
/clone_lib="Notre Dame Liverpool"
/notes="Vector: pECBAC1; Site_1: Hind III; The library was
prepared from whole body tissue of newly hatched L1 larvae
by David Severson at the University of Notre Dame and
Hongbin Zhang"

ORIGIN
Query Match 80.0%; Score 18.4; DB 28; Length 927;
Best Local Similarity 95.0%; Pred. No. 2.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATGAAGTACATACACACC 20
    |||||
Db 577 GAATGAAGTACATACACACC 596

RESULT 6
BE572261 999 bp mRNA linear EST 15-AUG-2000
601330149F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3707436 5',
mRNA sequence.
ACCESSION BE572261
VERSION BE572261.1 GI:9815981
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 999)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM8925 row: b column: 13
High quality sequence stop: 340.

FEATURES
source
LOCATION/Qualifiers
1..999
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3707436"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"

```

/note="Organ: mammary; Vector: pCW-SPORT6; Site: 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN

Query Match 80.0%; Score 18.4; DB 10; Length 999;
Best Local Similarity 95.0%; Pred. No. 2.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGAACCTACATACACACACC 23
||||| : |||||
Db 464 TGAACCTACATACACACACC 483

RESULT 7

AL572872/c 1201 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION AL572872 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
clone CSODI034YH01 3-PRIME, mRNA sequence.

ACCESSION
VERSION AL572872
KEYWORDS EST.

SOURCE
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

COMMENT On Feb 16, 2001 this sequence version replaced gi:12931557.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2214.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODI034CD01NP1&cluster=2214.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODI034CD01NP1.

FEATURES

source

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODI034YH01"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="First strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 80.0%; Score 18.4; DB 9; Length 1201;
Best Local Similarity 86.4%; Pred. No. 2.6e+03;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATACACACC 22
||||| : |||||
Db 526 GAATGAACCTACATACACACC 505

RESULT 8

AW698814/c 129 bp mRNA linear EST 17-APR-2000
LOCUS
DEFINITION AW698814 r406 non-glandular-haired subtracted cDNA library Medicago sativa
cDNA, mRNA sequence.

ACCESSION AW698814

VERSION
KEYWORDS
SOURCE
ORGANISM

AW698814.1 GI:7581400

EST.

Medicago sativa

Medicago sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

1 (bases 1 to 129)

Hays, D. and Skinner, D.

Expressed sequence tags subtracted in reciprocal fashion between glandular-haired and non-glandular-haired full sibs of alfalfa unpublished (2000)

JOURNAL

COMMENT

Contact: Hays DB

USDA, ARS, GMPRC, PSERU; Department of Agronomy

Kansas State University

Throckmorton Hall, Manhattan, KS 66506, USA

Tel: 785 532 7116

Fax: 785 532 6167

Email: dhays@genes.alfalfa.ksu.edu

Seq primer: SP6.

FEATURES

source

Location/Qualifiers

1..129

/organism="Medicago sativa"

/mol_type="mRNA"

/cultivar="Riley X KS224"

/db_xref="taxon:3879"

/tissue_type="Leaf and stem"

/clone_lib="non-glandular-haired subtracted cDNA library"

/note="glandular-haired versus non-glandular-haired

reciprocal cDNA subtraction with CLONTECH PCR-Select cDNA

subtraction."

ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 129;

Best Local Similarity 87.0%; Pred. No. 3.2e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATACACACC 23

||||| : |||||

Db 88 GAATGAACCTACATACACACC 66

RESULT 9

CD183415/c

LOCUS

DEFINITION

MS1-00380-A246-G09-U-B MS1-0038 Schistosoma mansoni cDNA clone

MS1-00380-A246-G09-B, mRNA sequence.

ACCESSION

CD183415

VERSION

CD183415.1 GI:34713619

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansoni

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 307)

Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,

Opoti, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,

Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,

Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,

Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,

Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,

Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,

Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,

Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

Transcriptome analysis of the acelomate human parasite Schistosoma

mansoni

JOURNAL

MEDLINE

COMMENT

Nat. Genet. 35 (2), 148-157 (2003)

22879926

Contact: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,

Brasil

Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjoeiq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: MS1-0038U-A246 row: 9 column: G.

FEATURES

Location/Qualifiers

1..307
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MS1-0038U-A246-G09.B"
/sex="mixed pool"
/dev_stage="schistosomulum"
/lab_host="in vitro culture"
/clone_lib="MS1-0038"
/note="Vector: pGEM T-easy"

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 307;

Best Local Similarity 87.0%; Pred. No. 3.1e+03;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAATGAACATACATAACACACC 23

|||||
Db 128 GAATAAATACATAACACACC 106

RESULT 10

CD133709

LOCUS

CD133709 309 bp mRNA linear EST 14-SEP-2003

MS1-0024U-A317-H10-U.G MGI-0024 Schistosoma mansoni cDNA clone

MS1-0024U-A317-H10.G, mRNA sequence.

ACCESSION

CD133709

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansoni

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 309)

Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,

Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,

Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,

Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,

Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,

Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,

Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,

Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,

Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

Transcriptome analysis of the acelomate human parasite Schistosoma

mansoni

Nat. Genet. 35 (2), 148-157 (2003)

22879926

CONTACT: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,

Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjoeiq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST

Genome Project. All sequences in the project were assembled and

annotated. This entry and all the assembled sequences can be seen

in the following URL <http://bioinfo.iq.usp.br/schisto/>

Plate: MGI-0024U-A317 row: 10 column: H.

FEATURES

Location/Qualifiers

1..309

/organism="Schistosoma mansoni"

/mol_type="mRNA"

/db_xref="taxon:6183"

/clone="MGI-0024U-A317-H10.G"

/sex="mixed pool"

/dev_stage="germball"

/lab_host="Biomphalaria glabrata"

/clone_lib="MGI-0024"

/note="Vector: pGEM T-easy"

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Qy

Db

RESULT 11

CD188719/c

LOCUS

CD188719

DEFINITION

MS1-0063U-A262-A04-U.B MS1-0063 Schistosoma mansoni cDNA clone

MS1-0063U-A262-A04.B, mRNA sequence.

CD188719

ACCESSION

CD188719

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansoni

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 313)

Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,

Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,

Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,

Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,

Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,

Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,

Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,

Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,

Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

Transcriptome analysis of the acelomate human parasite Schistosoma

mansoni

Nat. Genet. 35 (2), 148-157 (2003)

22879926

CONTACT: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,

Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjoeiq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST

Genome Project. All sequences in the project were assembled and

annotated. This entry and all the assembled sequences can be seen

in the following URL <http://bioinfo.iq.usp.br/schisto/>

Plate: MS1-0063U-A262 row: 4 column: A.

FEATURES

source

Location/Qualifiers

1..313

/organism="Schistosoma mansoni"

/mol_type="mRNA"

/db_xref="taxon:6183"

/clone="MS1-0063U-A262-A04.B"

/sex="mixed pool"

/dev_stage="schistosomulum"

/lab_host="in vitro culture"

/clone_lib="MS1-0063"

/note="Vector: pGEM T-easy"

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

79.1%; Score 18.2; DB 14; Length 313;

87.0%; Pred. No. 3.1e+03;

20; 0; 3; 0; 0;

Qy

Db

180 GAATAAATACATAACACACC 202

|||||

Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni
Nat. Genet. 35 (2), 148-157 (2003)
22879926

CONTACT: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil

Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL: <http://bioinfo.iq.usp.br/schisto/>
Plate: MA3-9999U-V243 row: 11 column: H.

FEATURES

Location/Qualifiers
1..537
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MA3-9999U-V243-H11.B"
/sex="mixed pool"
/dev_stage="adult"
/lab_host="Mus musculus"
/clone_lib="MA3-0001"

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 537;
Best Local Similarity 87.0%; Pred. No. 3.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACC 23
|||||
DB 460 GAATAAACTACATACACACC 438

RESULT 15

CD073220 556 bp mRNA linear EST 14-SEP-2003
MA3-0001U-L241-H07-U-B MA3-0001 Schistosoma mansoni cDNA clone
MA3-0001U-L241-H07.B, mRNA sequence.

CD073220
CD073220.1 GI:34624266

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 556)
Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

TITLE

JOURNAL

MEDLINE

COMMENT

Nat. Genet. 35 (2), 148-157 (2003)
22879926
CONTACT: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL: <http://bioinfo.iq.usp.br/schisto/>
Plate: MA3-0001U-L241 row: 7 column: H.

FEATURES

Location/Qualifiers
1..556
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MA3-0001U-L241-H07.B"
/sex="mixed pool"
/dev_stage="adult"
/lab_host="Mus musculus"
/clone_lib="MA3-0001"

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 556;
Best Local Similarity 87.0%; Pred. No. 3.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACC 23
|||||
DB 197 GAATAAACTACATACACACC 219

Search completed: March 25, 2004, 00:13:45
Job time : 1125.76 secs


```

REFERENCE
AUTHORS
TITLE
JOURNAL
12 (bases 1 to 4229)
Miyata,T., Nangaku,M., Inagi,R. and Kurokawa,K.
Direct Submission
Submitted (15-FEB-2000) Institute of Medical Sciences and
Department of Internal Medicine, Tokai University School of
Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
3 (bases 1 to 4229)
Miyata,T., Nangaku,M., Inagi,R. and Kurokawa,K.
Direct Submission
Submitted (29-MAR-2002) Institute of Medical Sciences and
Department of Internal Medicine, Tokai University School of
Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
REMARK
COMMENT
Sequence update by submitter
On Mar 29, 2002 this sequence version replaced gi:18000453.
FEATURES
Location/Qualifiers
source
1..4229
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21.3"
1..4021
promoter
4022..>4229
rRNA
/product="megsin"
ORIGIN
Query Match 93.6%; Score 23.4; DB 9; Length 4229;
Best Local Similarity 96.0%; Pred. No. 1.7;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AACCACTTAGTCAGACTACTT 25
Db 3936 AACCACTTAGTCAGACTACTT 3960

RESULT 2
AC072051/c
LOCUS AC072051 157284 bp DNA linear PRI 03-DEC-2001
DEFINITION Homo sapiens chromosome, clone RP11-79D21, complete sequence.
ACCESSION AC072051
VERSION AC072051.8 GI:16974280
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157284)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-79D21
Unpublished
2 (bases 1 to 157284)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Balding,J., Barna,N., Bastien,V., Bede,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lie,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., Mihova,T., McKernan,K., McPheters,R.,
Meldrim,J., Meneus,L., Mihoval,T., Miranda,C., Mlenaga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 157284)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Cooke,P., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fargo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Goyette,M., Graham,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenaga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 157284)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fargo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Goyette,M., Graham,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenaga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2001 this sequence version replaced gi:14277307.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L966
Center clone name: 79_D_21
-----
Location/Qualifiers
1..157284
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-79D21"
FEATURES
source

```


Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roberti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2002 this sequence version replaced gi:11990731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L8516
Center clone name: 317.G.1

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 85659: contig of 85659 bp in length
* 85660 85759: gap of 100 bp
* 85760 163014: contig of 77255 bp in length
* 163015 163114: gap of 100 bp
* 163115 189092: contig of 25978 bp in length.

FEATURES
source

1. 189092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-317G1"
/clone_lib="RP11-11 Human Male BAC"

ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 189092;
Best Local Similarity 96.0%; Pred. No. 1;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTACTTT 25
|||||
Db 66579 AACCACTTAGTCAGATACACTACTTT 66603

RESULT 4
AX346917/C
LOCUS
DEFINITION
ACCESSION

AX346917 6182 bp DNA linear PAT 01-FEB-2002
Sequence 1988 from Patent WO0200928.
AX346917

AX346917.1 GI:18494803

KEYWORDS

SOURCE
ORGANISM
synthetic construct
artificial sequences.

REFERENCE

1 Olek, A., Piepenbrock, C. and Berlin, K.
Diagnosis of diseases associated with the immune system
Patent: WO 0200928-A 1988 03-JAN-2002;
Epigenomics AG (DE)

FEATURES

source
Location/Qualifiers
1. 6182
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 6182;
Best Local Similarity 88.0%; Pred. No. 58;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTACTTT 25
|||||

Db 1295 AACCACTTAGTCAGATACACTACTTT 1271
|||||

RESULT 5

LOCUS

AC111981

DEFINITION

Rattus norvegicus clone CH210-121G8, WORKING DRAFT SEQUENCE, 5
unordered pieces.
AC111981

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 346601)
Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidaa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okwou, G., Olanpunsagoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfamknoch, C.,
 Plopper, F., Potndexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shactman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejo, Z., Usmani, K.,
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 346601)
 Worley, K.C.
 Direct Submission
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 346601)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23665344.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GOVY
 Center clone name: CH230-121G8
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 216460 bases at least Q40
 Consensus quality: 218712 bases at least Q30
 Consensus quality: 220592 bases at least Q20
 Estimated insert size: 221945; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 113609: contig of 113609 bp in length
 * 113610 113709: gap of unknown length
 * 113710 166592: contig of 52883 bp in length

RP11-12C17 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECIOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-12C17. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-12C17 is at 63011 in this sequence. The true left end of clone GSI-11815 is at 33117 in this sequence. The true right end of clone RP11-6713 is at 100 in this sequence.

FEATURES

source	Location/Qualifiers
repeat_region	1..63011
repeat_region	/organism="Homo sapiens"
repeat_region	/mol_type="genomic DNA"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="1"
repeat_region	/clone="RP11-12C17"
repeat_region	/clone_lib="RPCI-11.1"
repeat_region	1..238
repeat_region	/note="THE1C repeat: matches 1..243 of consensus"
repeat_region	312..598
repeat_region	/note="HERV23 repeat: matches 1..274 of consensus"
repeat_region	complement(576..1195)
repeat_region	/note="match: GSS: Em:AQ530043"
repeat_region	609..839
repeat_region	/note="LTR24 repeat: matches 250..480 of consensus"
repeat_region	1093..1195
repeat_region	/note="LTR24 repeat: matches 1..109 of consensus"
repeat_region	1796..1913
repeat_region	/note="L2 repeat: matches 2617..2735 of consensus"
repeat_region	2084..2295
repeat_region	/note="TIGER1 repeat: matches 863..1014 of consensus"
repeat_region	2388..2635
repeat_region	/note="MER44C repeat: matches 81..359 of consensus"
repeat_region	2636..3044
repeat_region	/note="L2 repeat: matches 2069..2441 of consensus"
repeat_region	3045..3405
repeat_region	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	3407..4975
repeat_region	/note="THE1B-INTERNAL repeat: matches 2..1580 of consensus"
repeat_region	4976..5343
repeat_region	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	5344..5423
repeat_region	/note="L2 repeat: matches 2441..2513 of consensus"
repeat_region	5424..5743
repeat_region	/note="Aluub repeat: matches 1..299 of consensus"
repeat_region	5744..5847
repeat_region	/note="L2 repeat: matches 2513..2722 of consensus"
repeat_region	7238..7379
repeat_region	/note="MIR repeat: matches 40..182 of consensus"
repeat_region	7683..7982
repeat_region	/note="Alusp repeat: matches 1..312 of consensus"
repeat_region	8098..8156
repeat_region	/note="MIR repeat: matches 121..182 of consensus"
repeat_region	8157..8456
repeat_region	/note="Alusx repeat: matches 1..297 of consensus"
repeat_region	8457..8574
repeat_region	/note="MIR repeat: matches 16..121 of consensus"
repeat_region	8804..9114
repeat_region	/note="Aluy repeat: matches 1..306 of consensus"
repeat_region	9593..9897
repeat_region	/note="Aluy repeat: matches 1..304 of consensus"
repeat_region	9898..10253
repeat_region	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	10470..10599
repeat_region	/note="FLAM C repeat: matches 1..132 of consensus"
repeat_region	10610..10903
repeat_region	/note="Alusg repeat: matches 1..293 of consensus"
repeat_region	10982..11180
repeat_region	/note="L2 repeat: matches 1986..2204 of consensus"
repeat_region	12003..12029
repeat_region	/note="MIR repeat: matches 79..102 of consensus"
repeat_region	12030..12761
repeat_region	/note="LTR8 repeat: matches 1..691 of consensus"
repeat_region	12762..12848
repeat_region	/note="MIR repeat: matches 102..205 of consensus"
repeat_region	12903..13455
repeat_region	/note="MER41A repeat: matches 1..554 of consensus"
repeat_region	13456..13540
repeat_region	/note="MER41-internal repeat: matches 1..86 of consensus"
repeat_region	complement(14302..14756)
repeat_region	/note="match: STS: Em:HSP03H12"
repeat_region	14302..14528
repeat_region	/note="MIR repeat: matches 3..259 of consensus"
repeat_region	14622..14743
repeat_region	/note="MIR repeat: matches 85..212 of consensus"
repeat_region	15285..15543
repeat_region	/note="L2 repeat: matches 2255..2518 of consensus"
repeat_region	15846..16227
repeat_region	/note="L1M4 repeat: matches 3865..4250 of consensus"
repeat_region	16251..16440
repeat_region	/note="MER20 repeat: matches 1..179 of consensus"
repeat_region	16446..16803
repeat_region	/note="MER51A repeat: matches 1..371 of consensus"
repeat_region	16804..18061
repeat_region	/note="MER51-internal repeat: matches 1..1250 of consensus"
repeat_region	complement(18014..18332)
repeat_region	/note="match: GSS: Em:AQ254526"
repeat_region	complement(18141..18319)
repeat_region	/note="match: GSS: Em:AQ023866"
repeat_region	18143..18323
repeat_region	/note="match: GSS: Em:AZ707492"
repeat_region	complement(18144..18333)
repeat_region	/note="match: GSS: Em:AZ468762"
repeat_region	18149..18192
repeat_region	/note="22 copies 2 mer at 75% conserved"
repeat_region	complement(18154..18330)
repeat_region	/note="match: GSS: Em:AQ417380"
repeat_region	18155..18333
repeat_region	/note="match: GSS: Em:AZ016479"
repeat_region	complement(18178..18350)
repeat_region	/note="match: GSS: Em:B73060"
repeat_region	complement(18181..18319)
repeat_region	/note="match: GSS: Em:B40172"
repeat_region	complement(18184..18340)
repeat_region	/note="match: GSS: Em:AQ683639"
repeat_region	complement(18187..18302)
repeat_region	/note="match: GSS: Em:AQ106777"
repeat_region	18265..18304
repeat_region	/note="10 copies 4 mer at 85% conserved"
repeat_region	18306..18333
repeat_region	/note="14 copies 2 mer at 96% conserved"
repeat_region	18354..18753
repeat_region	/note="LTR10C repeat: matches 188..580 of consensus"
repeat_region	18754..18829
repeat_region	/note="trna-Lys-AAG repeat: matches 1..76 of consensus"
repeat_region	18830..18860
repeat_region	/note="LTR10C repeat: matches 580..609 of consensus"
repeat_region	18861..18926
repeat_region	/note="MER51-internal repeat: matches 1186..1250 of consensus"
repeat_region	18928..19335
repeat_region	/note="MER51-internal repeat: matches 1181..1593 of consensus"
repeat_region	19336..19629
repeat_region	/note="MER42 repeat: matches 1..356 of consensus"
repeat_region	19630..20581
repeat_region	/note="MER51-internal repeat: matches 1593..2802 of consensus"
repeat_region	20582..20890
repeat_region	/note="AluSg1 repeat: matches 1..309 of consensus"
repeat_region	20891..21410
repeat_region	/note="MER51-internal repeat: matches 2802..3327 of

```

consensus"
repeat_region 21393..21893
/note="MER57-internal repeat: matches 3520..4039 of
consensus"
repeat_region 21873..21983
/note="MER61-internal repeat: matches 1584..1699 of
consensus"
repeat_region 21937..24469
/note="MER51-internal repeat: matches 5199..7816 of
consensus"
repeat_region 24471..24809
/note="MER51A repeat: matches 1..371 of consensus"
repeat_region 24852..25138
/note="AluJo repeat: matches 5..294 of consensus"
repeat_region 25144..25215
/note="LIM4 repeat: matches 3768..3836 of consensus"

Query Match 79.2%; Score 19.8; DB 9; Length 63011;
Best Local Similarity 91.3%; Pred. No. 66;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ACCACCTTACTCAGATGCTACTT 24
|||||
Db 31590 ACCACCTTACTCAGATGCTACTT 31568

RESULT 7
AC009946/c
LOCUS
DEFINITION Homo sapiens clone RP11-12C17, complete sequence.
AC009946
VERSION AC009946.2 GI:6604542
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 3 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Dec 20, 1999 this sequence version replaced gi:5836211.
CENTER project name: NH0012C17.
FEATURES
source
1..169072
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-12C17"

ORIGIN
Query Match 79.2%; Score 19.8; DB 9; Length 169072;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 ACCACCTTACTCAGATGCTACTT 24
|||||
Db 137651 ACCACCTTACTCAGATGCTACTT 137629

RESULT 8
AC139347
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP24-334F11, WORKING DRAFT
AC139347
ACCESSION AC139347.2 GI:28191622
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205202)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 205202)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 205202)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Feb 1, 2003 this sequence version replaced gi:28173257.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M BB0334F11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 201546 bases at least Q40
Consensus quality: 202453 bases at least Q30
Consensus quality: 202988 bases at least Q20
Insert size: 229000; agarose-fp
Quality coverage: 17.08 in Q20 bases; agarose-fp
Quality coverage: 11.09 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1135: contig of 1135 bp in length
* 1136 1235: gap of unknown length
* 1236 2878: contig of 1643 bp in length
* 2879 2978: gap of unknown length
* 2979 7471: contig of 4492 bp in length
* 7471 7570: gap of unknown length
* 7571 13052: contig of 5482 bp in length
* 13053 13153: gap of unknown length
* 13153 23285: contig of 10133 bp in length
* 23286 23385: gap of unknown length

```

* 23386 38432: contig of 15047 bp in length
 * 38433 38532: gap of unknown length
 * 38533 52633: contig of 14101 bp in length
 * 52634 52733: gap of unknown length
 * 52734 102557: contig of 49824 bp in length
 * 102558 102657: gap of unknown length
 * 102658 153484: contig of 50827 bp in length
 * 153485 153584: gap of unknown length
 * 153585 205202: contig of 51618 bp in length.

FEATURES

Location/Qualifiers
 1..205202

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP24-334F11"

misc_feature

1..1135
 /note="assembly_name:Contig8"

misc_feature

1236..2878
 /note="assembly_name:Contig10"

misc_feature

2979..7470
 /note="assembly_name:Contig11"

misc_feature

7571..13052
 /note="assembly_name:Contig12"

misc_feature

13153..23285
 /note="assembly_name:Contig13"

misc_feature

23386..38432
 /note="assembly_name:Contig14"

misc_feature

38533..52633
 /note="assembly_name:Contig15"

misc_feature

52734..102557
 /note="assembly_name:Contig16"

misc_feature

102658..153484
 /note="assembly_name:Contig17"

misc_feature

153585..205202
 /note="assembly_name:Contig18"

ORIGIN

Query Match 77.6%; Score 19.4; DB 2; Length 205202;
 Best Local Similarity 95.2%; Pred. No. 89;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACACCTTAGTCAGATACTA 21

Db 78563 AACACCTTAGTCAGATACTA 78583

RESULT 9

CLPA87KB

LOCUS CLPA87KB 8694 bp DNA linear PLN 29-MAR-2001
 DEFINITION Picea abies 8.7kb chloroplast DNA fragment.

ACCESSION

AJ001004

VERSION

AJ001004.1 GI:2764566

KEYWORDS

accD gene; acetyl-coenzyme A carboxylase; atpB gene; atpE gene;

H₂-ATP synthase; photosystem I polypeptide I; psal gene;

ribulose-bisphosphate carboxylase large subunit; trnA-Arg; trnR

gene.

SOURCE chloroplast Picea abies (Norway spruce)

ORGANISM Picea abies

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

1 Sutter, A., Philipps, A. and Wild, A.

Picea abies chloroplast genome fragment of 8,7 kb including atpE,

atpB, rbcL, trnR, accD, and psal

Unpublished

2 (bases 1 to 8694)

Sutter, A.

Direct Submission

Submitted (05-AUG-1997) Sutter A., Johannes Gutenberg-Universitaet

Mainz, Institut fuer Allgemeine Botanik, Mullerweg 6, 55099 Mainz,

GERMANY

Revised by author 24-NOV-97

FEATURES

source

Location/Qualifiers

1..8694

/organism="Picea abies"

/organelle="plastid;chloroplast"

/mol_type="genomic DNA"

/db_xref="taxon:3323"

/clone_lib="lambda ZAP II"

complement(234..647)

/gene="atpB"

complement(234..647)

/gene="atpB"

/codon_start=1

/transl_table=11

/product="H₂-ATP synthase subunit B"

/protein_id="CAA04457.1"

/db_xref="GI:2764567"

/db_xref="GOA:O47036"

/translation="MTLNRLVLSPNRVINDSEVKELILSTNSQIGVLNPHASIVAAV

DIGVMKIRLNGQSTWALMGGFPAKINDRITVLVNNARVDIDLKEAQETFKVAKAD

LARAEGKRAIEADVALKCKARTLEAISAPPVSN"

complement(656..2122)

/gene="atpB"

complement(656..2122)

/gene="atpB"

/codon_start=1

/transl_table=11

/product="H₂-ATP synthase subunit B"

/protein_id="CAA04458.1"

/db_xref="GI:2764568"

/db_xref="GOA:O47037"

/db_xref="SWISS-PROT:O47037"

/translation="MRINELVLGVLSALVKNQVRIAGIIPVLVDVFPFGNMPNINYS

LIVGQGTAGQEIVTCEVQQLGNHKAIVAMSDTGLTRGRMRVIDTGAPLSPVYGG

ATLGRFNVLGEFVDNLGPVDARITSPHRSAPAFTELDTKLSIFETGKVVLLAPY

RRGKICLFGAGVGKTVLIMELINNAKAHGVSVFGVGRTERGNDLYMEMKESG

VIDEONISKVALVYQMNPEPGARMVGLTALTWAEYFRDVNEODVLSFDNIIFRF

VQAGSEVSALLGRMPSAVCYOPTLATMSGLOERITSTKRGSTITSQAVYVADDLTD

PAPATFHLDATTVPFSLAAGIYPVDPDLDSTMLQPMIVGEERTYETAQGVKQT

LQRYKELQDIILAIPLGLDEUSEEDRLIVARAKIERFLSQFFFFVAVFTGPGKYVGLM

ETIRGFQMLTSGLEQSFYLVGNDEATAKANMSKTES"

4172..5599

/gene="rbcL"

4172..5599

/gene="rbcL"

/codon_start=1

/transl_table=11

/product="ribulose-bisphosphate carboxylase large subunit"

/protein_id="CAA04459.1"

/db_xref="GI:2764569"

/db_xref="GOA:P48711"

/db_xref="SWISS-PROT:P48711"

/translation="NSPKTETKASVGFAGVKDYRLTYTYTPEYQTKDIDLAAFRVTP

QFQVPEEAGAAVAASSTGTWTVTWDTGLTSLDRYKGRCYDIEPVAGEESQFIAPVA

YPLDFEESGVNTLFTSIVGNVFGKALRALRLEDLRIPPAYSKTFQGPFGHGIQVERD

KINKYRPLGCTIKPKLGSANKYGRAVVECLRGLOFTKDDENVNSQPFWRDRF

VFCALYKAQAFTEIKGHYLNATAGTCEEMKRAVFAVELGPIVMHDYLTGGFTA

NTSLAHYCRDNGLLHHRAMHAVIDRQKNHGMHPRVLAKALRMGGDHIGGTVVGK

LEGEREITLGFVLLDRDFIEKDRSGIYFTQDMVMPGLVPSVSGGIHVMHPALTE

IFGDDSVLQFGGGTGLHFWGNAPGAVANRVALEACVQARNEGRDILARENEVIREASK

WSPELAAACEIKWEIKFPEAVDTI"

5915..5988

/gene="trnR"

5915..5988

/gene="trnR"

/product="trnA-Arg"

6244..7203

/gene="accD"

6244..7203

/gene="accD"

/codon_start=1

/transl_table=11

/product="acetyl-coenzyme A decarboxylase"

```

/protein_id="CAA04460.1"
/db_xref="GI:2764570"
/db_xref="GOA:047039"
/db_xref="SWISS-PROT:047039"
/translation="MSIRWFEDRRKITGLKNSVERDSKOVNERNKNLSIDYVKI
NRLVQCNCESLLYIRFRENKSCVECYLQWNSDRIELLIDRGTRHPMEDMY
TLIDVQFHSNENPAHSDPLHSDESKDHITFCQIETGLTDAIQTGOLNGLPIALG
VNDKFMGSMGVSVEKHTLRIERATESLPVIMVCASGGARMORGSSFSMOMAKIA
SALYIPQKDNRLLYVSLTSPPTGGTASFGMLGDLIIARPKAYTAFAGKIVIDQTLG
QKVIDFQVIEHLFGHGLDLIVRNLLKCVLSLFIYVLQSS"
7765..7875
/gene="psai"
7765..7875
/gene="psai"
/codon_start=1
/transl_table=11
/product="Photosystem I polypeptide I"
/protein_id="CAA04461.1"
/db_xref="GI:2764571"
/db_xref="GOA:047040"
/db_xref="SWISS-PROT:047040"
/translation="MIFNLPSEFFVLVGLLPAITWVIFHLIYIQNDIF"

ORIGIN
Query Match 76.8%; Score 19.2; DB 8; Length 8694;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACACCTTAGTCAGTACTACTTT 25
Db 2250 ACCACCTTAGTAATCTACTTT 2273
|||||
|||||

RESULT 10
AC011752
LOCUS Homo sapiens BAC clone RP11-531P14 from 2, complete sequence.
DEFINITION AC011752
ACCESSION AC011752.2 GI:7534296
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Suleston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 121227)
Tonn,M., Drone,K. and Scott,K.
The sequence of Homo sapiens BAC clone RP11-531P14
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 121227)
Waterston,R.H.
Direct Submission
JOURNAL
REFERENCE 4 (bases 1 to 121227)
Submitted (13-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
JOURNAL
REFERENCE 5 (bases 1 to 121227)
Waterston,R.H.
Direct Submission
JOURNAL
REFERENCE 6 (bases 1 to 121227)
Submitted (11-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
JOURNAL
REFERENCE 7 (bases 1 to 121227)
Waterston,R.
Direct Submission
JOURNAL
REFERENCE 8 (bases 1 to 121227)
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 11, 2000 this sequence version replaced gi:6042116.
----- Genome Center

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapien@wustl.edu
----- Summary Statistics
Center project name: H_NH0531P14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-500G18, 200 base pair overlap. Actual start of this clone is at base position 1 of RP11-531P14; actual end is at base position 34258 of RP11-500G18.

RP11-531P14 contains a transposon in the growth of the clone, which is not part of the submitted sequence.

FEATURES	Location/Qualifiers
source	1..121227
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-531P14"
	/clone_lib="RPCI-11"
	1805..2244
	/rpt_family="L2"
	2436..2451
	/note="similar to EST AA648276 (NID:g2574705) ns20d03.s1"
	3307..3349
	/rpt_family="L2"
	4249..4270
	/rpt_family="AT_rich"
	5090..5111
	/rpt_family="(TG)n"
	5323..5538
	/rpt_family="L2"
	6651..6711
	/rpt_family="Ricksha"
	6712..6949
	/rpt_family="MER2_type"
	6965..7417
	/rpt_family="Retroviral"
	8545..8614
	/rpt_family="MIR"

```

repeat_region 9587. .9635
/rpt_family="L2"
repeat_region 9655. .10026
/rpt_family="L2"
repeat_region 10057. .10199
/rpt_family="L2"
repeat_region 11189. .11353
/rpt_family="L2"
repeat_region 11388. .11793
/rpt_family="L2"
repeat_region 11792. .12080
/rpt_family="L2"
repeat_region 12102. .12297
/rpt_family="L2"
repeat_region 12687. .12790
/rpt_family="L2"
repeat_region 12790. .12843
/rpt_family="L2"
repeat_region 14002. .14282
/rpt_family="Retroviral"
repeat_region 14387. .15337
/rpt_family="Retroviral"
repeat_region 15329. .15381
/rpt_family="BUR1"
repeat_region 15403. .16925
/rpt_family="MER1_type"
repeat_region 16949. .17073
/rpt_family="Alu"
repeat_region 17074. .17929
/rpt_family="MER1_type"
repeat_region 17930. .18222
/rpt_family="Alu"
misc_feature 17930. .17943
/notes="similar to EST AA554531 (NID:G2325070) nl14a02.s1"
repeat_region 18223. .18354
/rpt_family="MER1_type"
repeat_region 18378. .19486
/rpt_family="Retroviral"
repeat_region 19521. .19992
/rpt_family="Retroviral"
repeat_region 20171. .20317
/rpt_family="L2"
repeat_region 20318. .20610
/rpt_family="Alu"
repeat_region 20611. .20656
/rpt_family="L2"
repeat_region 23453. .23586
/rpt_family="MER1_type"
repeat_region 23608. .23783
/rpt_family="MIR"
repeat_region 25428. .25451
/rpt_family="AT-rich"
repeat_region 25484. .25599
/rpt_family="L2"
repeat_region 25974. .26000
/rpt_family="L2"
repeat_region 26101. .26155
/rpt_family="L2"
misc_feature 26589. .26936
/notes="match to EST AI081859 (NID:G3418651) ov24c04.x1"
misc_feature 27274. .27725
/notes="similar to EST AI954720 (NID:G5747030) wq32d06.x1"
repeat_region 28216. .28423
/rpt_family="L2"
repeat_region 28540. .28617
/rpt_family="L2"
repeat_region 29020. .29096
/rpt_family="MER2_type"
repeat_region 29097. .29622
/rpt_family="L1"
repeat_region 29645. .30187
/rpt_family="L1"
repeat_region 30263. .30288

```

```

/rpt_family="Alu"
30439. .30706
/rpt_family="Alu"
30722. .30879
/rpt_family="L1"
30967. .31089
/rpt_family="L1"
31097. .31555
/rpt_family="L1"
31555. .31840
/rpt_family="L1"
33392. .33438
/rpt_family="L2"
33730. .33784
/rpt_family="L2"
33821. .34119

Query Match 76.8%; Score 19.2; DB 9; Length 121227;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCACCTTAGTCAGATACACTACTTT 25
||||| ||||| ||||| ||||| |||||
Db 64479 ACCACATTAGTCAATCCTACTTT 64502

RESULT 11
AC131675 191085 bp DNA linear ROD 11-NOV-2003
LOCUS Mus musculus BAC clone RP23-328L1 from chromosome 18, complete
DEFINITION sequence.
ACCESSION AC131675
VERSION AC131675.4 GI:29825883
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Isak, A. and Bielicki, L.
1 (bases 1 to 191085)
The sequence of Mus musculus BAC clone RP23-328L1
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 191085)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 191085)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 191085)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 191085)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 191085)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 14, 2003 this sequence version replaced gi:29244816.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics

```

Center project name: M_BA0328L01

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tatenko in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC114820 and AC121874.

FEATURES

source

Location/Qualifiers

1. .131085

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"
/clone="RP23-328L1"

/clone_lib="RPCI-23"
1033. .1180

/rpt_family="Alu"
1451. .1655

/rpt_family="B2"
1908. .1990

/rpt_family="L1"
2060. .2102

/rpt_family="B4"
2073. .2211

/rpt_family="B2"
3367. .3498

/rpt_family="Alu"
3509. .3651

/rpt_family="Alu"
3694. .3750

/rpt_family="L1"
3760. .3838

/rpt_family="B4"
3839. .3997

/rpt_family="B4"
4013. .4179

/rpt_family="B2"
4316. .4473

/rpt_family="B4"
4685. .4787

/rpt_family="L1"
4796. .4930

/rpt_family="Alu"
5004. .5288

/rpt_family="L1"
5353. .5460

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region
5631. .5766
/rpt_family="Alu"
repeat_region
5939. .6087
/rpt_family="B4"
repeat_region
6039. .6268
/rpt_family="Alu"
repeat_region
6667. .6844
/rpt_family="B2"
repeat_region
6887. .7065
/rpt_family="B2"
repeat_region
7076. .7137
/rpt_family="B2"
repeat_region
7216. .7327
/rpt_family="B2"
repeat_region
7442. .7561
/rpt_family="Alu"
repeat_region
7563. .7703
/rpt_family="L1"
repeat_region
7738. .7924
/rpt_family="Alu"
repeat_region
7973. .8061
/rpt_family="B2"
repeat_region
8076. .8184
/rpt_family="Alu"
repeat_region
8186. .8295
/rpt_family="L1"
repeat_region
8347. .8469
/rpt_family="Alu"
repeat_region
8875. .9053
/rpt_family="B2"
repeat_region
9336. .9429
/rpt_family="B2"
repeat_region
9459. .9577
/rpt_family="B2"
repeat_region
9761. .9860
/rpt_family="L1"
repeat_region
9866. .10001
/rpt_family="L1"
repeat_region
10221. .10442
/rpt_family="B2"
repeat_region
10454. .10576
/rpt_family="B4"
repeat_region
10625. .10705
/rpt_family="B4"
repeat_region
10903. .11053
/rpt_family="Alu"
repeat_region
11179. .11299
/rpt_family="B4"
repeat_region
11333. .11497
/rpt_family="L1"
repeat_region
11556. .11661
/rpt_family="B2"
repeat_region
11674. .11806
/rpt_family="B2"
repeat_region
12002. .12078
/rpt_family="Alu"
repeat_region
12010. .12098
/rpt_family="B4"
repeat_region
13196. .13267
/rpt_family="Alu"
repeat_region
13626. .13726
/rpt_family="ID"
repeat_region
14236. .14307
/rpt_family="Alu"
repeat_region
14441. .15450
/rpt_family="Alu"
misc_feature
/note="CpG_island (%GC=72.3, o/e=0.94, #CpGs=114)"
15676. .15749
/rpt_family="B4"
repeat_region
15851. .15949
/rpt_family="B4"
repeat_region
15985. .16138
/rpt_family="B4"

repeat_region 16218..16285
 /rpt_family="ID"
 repeat_region 16472..16523
 /rpt_family="Alu"
 repeat_region 16697..16776
 /rpt_family="Alu"
 repeat_region 16779..16923
 /rpt_family="B4"
 repeat_region 16977..17064

Query Match 76.8%; Score 19.2; DB 10; Length 191085;
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACTACTT 24

Db 12117 AACACGTTACTCAGATACTACTT 12140

RESULT 12
 AL807818/c 194425 bp DNA linear VRT 18-FEB-2003
 DEFINITION Zebrafish DNA sequence from clone CH211-225H24, complete sequence.
 ACCESSION AL807818
 VERSION AL807818.14 GI:28412565
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 Beasley, H.
 Direct Submission

Submitted (18-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Feb 18, 2003 this sequence version replaced gi:28273001.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

COMMENT

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information
 on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
 beginning 'Dr' were identified by the Recon repeat discovery system
 (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
 were identified by Rick Waterman (Stephen Johnson lab, WashU). For
 further information see http://www/Projects/D_rerio/fishmask.shtml
 CH211-225H24 is from a CHORI-211 BAC library
 VECTOR: PTARBAC2.1.

FEATURES

Location/Qualifiers
 1..194425
 /organism="Danio rerio"
 /mol_type="genomic DNA"

/db_xref="taxon:7955"
 /clone="CH211-225H24"
 /clone_lib="CHORI-211"

Query Match 76.8%; Score 19.2; DB 5; Length 194425;
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACTACTT 24

Db 165858 AACACACTTACTCAGATACTACTT 165835

RESULT 13
 AC129307/c

LOCUS
 DEFINITION 198602 bp DNA linear HTG 25-AUG-2002
 Mus musculus chromosome UNK clone RP24-200K10, WORKING DRAFT
 SEQUENCE, 15 unordered pieces.

ACCESSION AC129307
 VERSION AC129307.3 GI:22476079
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Mus musculus (house mouse)

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone

Unpublished
 2 (bases 1 to 198602)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 1 (bases 1 to 198602)
 Submitted (28-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

3 (bases 1 to 198602)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 1 (bases 1 to 198602)
 Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

On Aug 25, 2002 this sequence version replaced gi:22138611.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUDSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@wustl.edu

----- Project Information -----
 Center project name: M.BB0200K10

----- Summary Statistics -----
 Sequencing vector: M13; 0%

Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319
 Consensus quality: 193934 bases at least Q40

Consensus quality: 195486 bases at least Q30
 Consensus quality: 196639 bases at least Q20

Insert size: 202000; agarose-fp
 Insert size: 197629; sum-of-contigs

Quality coverage: 12.83 in Q20 bases; agarose-fp
 Quality coverage: 11.64 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1099: contig of 1099 bp in length

```

* 1100 1199: gap of unknown length
* 1200 2248: contig of 1049 bp in length
* 2249 2348: gap of unknown length
* 2349 3711: contig of 1363 bp in length
* 3712 3811: gap of unknown length
* 3812 5920: contig of 2109 bp in length
* 5921 6020: gap of unknown length
* 6021 7194: contig of 1174 bp in length
* 7195 7294: gap of unknown length
* 7295 8673: contig of 1379 bp in length
* 8674 14281: contig of 5508 bp in length
* 14282 26301: contig of 11920 bp in length
* 26302 26401: gap of unknown length
* 26402 34164: contig of 7663 bp in length
* 34165 47001: contig of 12837 bp in length
* 47002 65387: contig of 18286 bp in length
* 65388 65487: gap of unknown length
* 65488 86634: contig of 21147 bp in length
* 86635 129492: contig of 42758 bp in length
* 129493 129592: gap of unknown length
* 129593 197861: contig of 68269 bp in length
* 197862 197962: contig of 641 bp in length.

```

FEATURES

source

```

1. .198602
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="UNK"
   /clone="RP24-200K10"

```

```

misc_feature 1. .1099
              /note="assembly_name:Contig39"
misc_feature 1200..2248
              /note="assembly_name:Contig45"
misc_feature 2349..3711
              /note="assembly_name:Contig48"
misc_feature 3812..5920
              /note="assembly_name:Contig49"
misc_feature 6021..7194
              /note="assembly_name:Contig50"
misc_feature 7295..8673
              /note="assembly_name:Contig51"
misc_feature 8774..14281
              /note="assembly_name:Contig52"
misc_feature 14382..26301
              /note="assembly_name:Contig53"
misc_feature 26402..34084
              /note="assembly_name:Contig54"
misc_feature 34165..47001
              /note="assembly_name:Contig55
              clone_end:SP6
              vector_side:left"
misc_feature 47102..65387
              /note="assembly_name:Contig56"
misc_feature 65488..86634
              /note="assembly_name:Contig57"
misc_feature 86735..129492
              /note="assembly_name:Contig58"
misc_feature 129593..197861
              /note="assembly_name:Contig59"
misc_feature 197962..198602
              /note="assembly_name:Contig31"

```

ORIGIN

```

Query Match      76.8%; Score 19.2; DB 2; Length 198602;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AACCACTTACTGTCAGATACTACTT 24
   |||||
Db 78202 AACACATTAGTCAGATACTAATT 78179

```

RESULT 14

AC122243/c

```

LOCUS AC122243 204584 bp DNA linear ROD 04-NOV-2003
DEFINITION Mus musculus chromosome 17 clone RP23-148C10, complete sequence.
ACCESSION AC122243
VERSION AC122243.3 GI:38154054
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE 1 (bases 1 to 204584)
AUTHORS Wilson,R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 204584)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 204584)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 204584)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 4, 2003 this sequence version replaced gi:34495085.

```

FEATURES

source

```

1. .204584
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="17"
   /clone="RP23-148C10"

```

ORIGIN

```

Query Match      76.8%; Score 19.2; DB 10; Length 204584;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AACCACTTACTGTCAGATACTACTT 24
   |||||
Db 143956 AACCAACAAGTCAGATACTACTT 143933

```

RESULT 15

AC100052/c

```

LOCUS AC100052 255457 bp DNA linear HTG 12-DEC-2003
DEFINITION Mus musculus chromosome 8 clone RP23-32L24 map 8, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AC100052
VERSION AC100052.6 GI:39752800
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```


Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 255457)

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Birken, B., Linton, L., Nusbäum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (bases 1 to 255457)

Birken, B., Nusbäum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, P., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (12-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Dec 12, 2003 this sequence version replaced gi:38454416. All repeats were identified using RepeatMasker: Smit, A. F. A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIPR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L13875

Center clone name: 32_L_24

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 101651: contig of 101651 bp in length
* 101652 101751: gap of 100 bp
* 101752 112879: contig of 11128 bp in length
* 112880 112979: gap of 100 bp
* 112980 204131: contig of 91152 bp in length
* 204132 233161: gap of 100 bp
* 233162 233261: contig of 28930 bp in length
* 233262 240236: gap of 100 bp
* 240237 240397: contig of 7035 bp in length
* 240397 255457: gap of 100 bp
* 255457: contig of 15061 bp in length.

FEATURES

source

1..255457
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clone="RP23-32L24"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 255457;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTGTAGTCAGATACCTT 24

DB 93399 AACACATTAGTCAGATACCTT 93376

Search completed: March 24, 2004, 23:01:26

Job time : 557.809 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 18:56:32 ; Search time 130.899 Seconds

(without alignments)
811.351 Million cell updates/sec

Title: US-09-889-611A-20

Perfect score: 25

Sequence: 1 aaccaccttagtcagataactattt 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAA71453	Aaa71453 Human meg
2	25	100.0	128	AAA71434	Aaa71434 Human meg
3	25	100.0	1431	AAA71435	Aaa71435 Human meg
4	22	88.0	30	AAA71449	Aaa71449 Human meg
c 5	20.2	80.8	6182	ABL34015	Ab134015 Human inn
c 6	18.8	75.2	110000	2	Continuation (17 o
7	18.6	74.4	471	AAK80155	Aak80155 Human inn
8	18.6	74.4	471	AAK80158	Aak80158 Human inn
9	18.6	74.4	471	AAK80159	Aak80159 Human inn
c 10	18.6	74.4	1523	3	AAC39949 Arabidops
11	18.6	74.4	2000	7	ADA73153 Rice gene
12	18.6	74.4	2999	9	ACF79501 Mouse pla
13	18.6	74.4	2999	9	ACF79502 Mouse pla
14	18.6	74.4	3053	7	ABF41815
15	18.6	74.4	3053	9	ACF79503 Rat plas
c 16	18.6	74.4	8058	6	ABN95829 Gene #232
c 17	18.6	74.4	24142	6	Abg76613 C. albica
c 18	17.8	71.2	28564	9	ADd46508 Human gen
c 19	17.6	70.4	369	3	AAC28947 Human sec
c 20	17.6	70.4	412	6	ABL68782 Kidney ca
21	17.6	70.4	803	4	Aai34790 Human neu
c 22	17.6	70.4	989	4	Aah01963 Candida i
c 23	17.6	70.4	1074	4	Aah01569 Candida d

c 24	17.6	70.4	1637	4	AAK69081	Aak69081 Human inn
25	17.6	70.4	1721	2	AAQ14626	Aaq14626 Human gli
26	17.6	70.4	1721	2	AAT34366	Aat34366 Plasmid p
27	17.6	70.4	1721	2	AAZ32337	Aaz32337 Human gli
28	17.6	70.4	1721	3	AAA88172	Aaa88172 pUC99 hum
29	17.6	70.4	2435	6	ABS68447	Abs68447 DNA encod
c 30	17.6	70.4	2790	4	AAK69082	Aak69082 Human inn
c 31	17.6	70.4	4708	1	AAN50415	Aan50415 Human ace
c 32	17.6	70.4	6071	9	ADB58365	Adb58365 Toxicity-
c 33	17.6	70.4	6071	9	ADB52939	Adb52939 Primary r
c 34	17.4	69.6	2967	4	AAS56729	Aas56729 Human BRC
35	17.2	68.8	5149	2	AAT32455	Aat32455 Galpain 1
36	17.2	68.8	30967	2	AAT32454	Aat32454 Galpain 1
37	17	68.0	480	7	ACF70412	Acf70412 Photorhab
38	17	68.0	694	7	ACF66756	Acf66756 Photorhab
39	17	68.0	762	7	ACF66728	Acf66728 Photorhab
c 40	17	68.0	762	7	ABX06371	Abx06371 S. pneumo
c 41	17	68.0	777	7	ABZ42202	Abz42202 Streptoco
c 42	17	68.0	1435	9	ADB58387	Adb58387 Toxicity-
c 43	17	68.0	1435	9	ADB52966	Adb52966 Primary r
c 44	17	68.0	2516	2	AAV52381	Aav52381 Streptoco
c 45	17	68.0	3191	4	AAF29742	Aaf29742 Castor be

ALIGNMENTS

RESULT 1
AAA71453
ID AAA71453 standard; DNA; 25 BP.
XX
AC AAA71453;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megain promoter PCR primer SEQ ID NO: 20.
XX
KW Promoter; megain; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP0000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
(MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megain useful for screening proteins.
XX
PS Example 4; Page 42; 45pb; Japanese.
XX

This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the invention

Sequence 25 BP; 8 A; 7 C; 2 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 3; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.057;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGAGATACTACTTT 25
Db 1 AACCACTTACTGAGATACTACTTT 25

RESULT 2

AAA71434
ID AAA71434 standard; DNA; 128 BP.

AC AAA71434;

XX 01-DEC-2000 (first entry)

XX Human megasin promoter fragment DNA.

XX Promoter; megasin; human; protein isolation; screening. ss.

XX Homo sapiens.

XX WO200043528-A1.

XX 27-JUL-2000.

XX 25-JAN-2000; 2000WO-JP000350.

XX 25-JAN-1999; 99JP-00015667.

XX (KURO/) KUROKAWA K.

XX (MIYA/) MIYATA T.

XX Miyata T;

XX WPI; 2000-543257/49.

XX DNA for promoter region of megasin useful for screening proteins.

XX Claim 1; Page 32; 45pp; Japanese.

XX This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents the human megasin promoter which is described in the method of the invention

XX Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;

Query Match 100.0%; Score 25; DB 3; Length 128;

Best Local Similarity 100.0%; Pred. No. 0.07; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGAGATACTACTTT 25

Db 16 AACCACTTACTGAGATACTACTTT 40

RESULT 3

AAA71435

ID AAA71435 standard; DNA; 1431 BP.

AC AAA71435;

XX 01-DEC-2000 (first entry)

XX Human megasin promoter fragment DNA.

XX Promoter; megasin; human; protein isolation; screening. ss.

XX Homo sapiens.

XX WO200043528-A1.

XX

XX 27-JUL-2000.

XX 25-JAN-2000; 2000WO-JP000350.

XX 25-JAN-1999; 99JP-00015667.

XX (KURO/) KUROKAWA K.

XX (MIYA/) MIYATA T.

XX Miyata T;

XX WPI; 2000-543257/49.

XX DNA for promoter region of megasin useful for screening proteins.

XX Disclosure; Fig 2; 45pp; Japanese.

XX This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents a fragment of the human megasin promoter which is described in the method of the invention

XX Sequence 1431 BP; 466 A; 266 C; 303 G; 394 T; 0 U; 2 Other;

Query Match 100.0%; Score 25; DB 3; Length 1431;

Best Local Similarity 100.0%; Pred. No. 0.095;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGAGATACTACTTT 25

Db 1319 AACCACTTACTGAGATACTACTTT 1343

RESULT 4

AAA71449

ID AAA71449 standard; DNA; 30 BP.

AC AAA71449;

XX 01-DEC-2000 (first entry)

XX Human megasin promoter PCR primer SEQ ID NO: 16.

XX Promoter; megasin; human; protein isolation; screening. PCR primer; ss.

XX Homo sapiens.

XX WO200043528-A1.

XX 27-JUL-2000.

XX 25-JAN-2000; 2000WO-JP000350.

XX 25-JAN-1999; 99JP-00015667.

XX (KURO/) KUROKAWA K.

XX (MIYA/) MIYATA T.

XX Miyata T;

XX WPI; 2000-543257/49.

XX DNA for promoter region of megasin useful for screening proteins.

XX Example 5; Page 40; 45pp; Japanese.

XX This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the

CC above vector; and (3) protein produced using (1). (1) is useful for
 CC screening and isolating proteins (especially transcription factors).
 CC AAA71434-A71469 represent PCR primers used in the method described in the
 CC invention

SQ Sequence 30 BP; 12 A; 9 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 88.0%; Score 22; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTAC 22
 DB 9 AACCCCTTAGTCAGATACACTAC 30

RESULT 5

ABL34015/c
 ID ABL34015 standard; DNA; 6182 BP.

AC ABL34015;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1988.

KW Human; immune system disease; cytosine methylation; antiaesthatic;
 KW antiarteriosclerotic; anianaemic; cytosine; neoptotic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DB-01032529.

PR 01-SEP-2000; 2000DB-01043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

SQ Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 6182;

Best Local Similarity 88.0%; Pred. No. 19;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTACTTT 25
 DB 1295 AACCCCTTAGTCAGATACACTACTTT 1271

RESULT 6

AAT42063_16/c

Continuation (17 of 19) of AAT42063 from base 1600001 (Haemophilus influenzae complete g
 WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063

WP Fragment Name Begin End

WP AAT42063_00 1 110000
 WP AAT42063_01 100001 210000
 WP AAT42063_02 200001 310000
 WP AAT42063_03 300001 410000
 WP AAT42063_04 400001 510000
 WP AAT42063_05 500001 610000
 WP AAT42063_06 600001 710000
 WP AAT42063_07 700001 810000
 WP AAT42063_08 800001 910000
 WP AAT42063_09 900001 1010000
 WP AAT42063_10 1000001 1110000
 WP AAT42063_11 1100001 1210000
 WP AAT42063_12 1200001 1310000
 WP AAT42063_13 1300001 1410000
 WP AAT42063_14 1400001 1510000
 WP AAT42063_15 1500001 1610000
 WP AAT42063_16 1600001 1710000
 WP AAT42063_17 1700001 1810000
 WP AAT42063_18 1800001 1830121

Query Match 75.2%; Score 18.8; DB 2; Length 110000;

Best Local Similarity 90.9%; Pred. No. 1.2e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTAC 22

DB 11158 AACCCCTTAGTCAGATACACTAC 11137

RESULT 7

AAK80155

ID AAK80155 standard; DNA; 471 BP.

AC AAK80155;

XX 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34967.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytosine; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

OS WO200157182-A2.

PN 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0198974P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214866P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218230P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226686P.
 PR 23-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX Disclosure; SEQ ID NO 34967; 3071pp + Sequence Listing; English.

XX AA454951 to AA464702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AA462170 to AA461921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the

CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
XX
SQ Sequence 471 BP; 157 A; 69 C; 118 G; 127 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 4; Length 471;
Best Local Similarity 84.0%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGACATCTACTTT 25
Db 383 AAGGACCTTGGCAGATCTACTTT 407

RESULT 8

AAK80158
ID AAK80158 standard; DNA; 471 BP.

XX
AC AAK80158;

XX
DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34970.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.

XX
XX W0200157182-A2.

XX
PD 09-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US001354.

XX
XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225367P.

PR 14-AUG-2000; 2000US-0225368P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 18-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226688P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-024523P.
PR 08-NOV-2000; 2000US-024524P.
PR 08-NOV-2000; 2000US-024525P.
PR 08-NOV-2000; 2000US-024526P.
PR 08-NOV-2000; 2000US-024527P.
PR 08-NOV-2000; 2000US-024528P.
PR 08-NOV-2000; 2000US-024532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.

PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239933P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.

PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 DR
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Disclosure; SEQ ID NO 34971; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 471 BP; 157 A; 69 C; 118 G; 127 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 4; Length 471;
 Best Local Similarity 84.0%; Pred. No. 78;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 AACCCCTTAGTCAGATACACTTTT 25
 Db 383 AAGGACCTTGGCAGATACACTTTT 407
 RESULT 10
 AAC39949/c
 ID AAC39949 standard; DNA; 1523 BP.
 XX
 AC AAC39949;
 XX
 XX 17-OCT-2000 (first entry)
 DT
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 26493.
 DE
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 XX
 XX EP1033405-A2.
 PN
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-00301439.
 PF
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.


```

PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 74.4%; Score 18.6; DB 3; Length 1523;
Best Local Similarity 84.0%; Pred. No. 90;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTCTTT 25
Db 645 AACCCCTTAGTCAGATACACTCTTT 621

RESULT 11
ADA73153
ID ADA73153 standard; DNA; 2000 BP.
AC ADA73153;
XX
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 6479.
XX
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; db.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to

```

```

PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
XX gene expression.
PS Claim 27; SEQ ID NO 6479; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 475 A; 562 C; 443 G; 519 T; 0 U; 1 Other;

Query Match 74.4%; Score 18.6; DB 7; Length 2000;
Best Local Similarity 84.0%; Pred. No. 93;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTCTTT 25
Db 403 AACCCCTTAGTCAGATACACTCTTT 427

RESULT 12
ACF79501
ID ACF79501 standard; cDNA; 2999 BP.
XX
XX ACF79501;
XX
XX 18-DEC-2003 (first entry)
XX
DE Mouse plasminogen activator inhibitor-1 cDNA.
XX
XX Plasminogen activator inhibitor-1; PAI-1; mouse; transgenic;
KW thrombolytic; antiasthmatic; antiinflammatory; nootropic;
KW neuroprotective; antidepressant; nephrotropic; vulnerary;
KW antiseborrheic; dermatological; antiarteriosclerotic; hepatotropic;
KW gene; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 132..1340
FT /*tag= a
FT /product= "Mouse PAI-1"
FT
XX
PN WO2003071267-A1.
XX
XX 28-AUG-2003.
XX
PF 19-FEB-2003; 2003WO-US005008.
XX
PR 19-FEB-2002; 2002US-0358061P.
XX
PA (UYVA-) UNIV VANDERBILT.
PA (DECL/) DECLERCK P J.
XX
PI Declerck PJ, Vaughan DE, Eren M;
XX
DR WPI; 2003-721694/68.
DR P-PSDB; A863126.
XX
XX
PT Treating a warm-blooded vertebrate animal having a medical condition,
e.g. Alzheimer's disease, atherosclerosis, myelofibrosis or
PT glomerulosclerosis, comprises administering a plasminogen activator
inhibitor-1 inhibitor.
XX
XX Disclosure; Page 73-76; 91pp; English.
PS

```

XX CC The present sequence is that of cDNA encoding mouse plasminogen activator
 CC inhibitor-1 (PAI-1). The PAI-1 gene can be used to produce transgenic
 CC animals of the invention, useful for screening potential PAI-1
 CC inhibitors. Such PAI-1 inhibitors can be used to treat a warm-blooded
 CC vertebrate animal having a medical condition, e.g. alopecia, undesired
 CC weight loss, Alzheimer's disease, systemic amyloidosis, myelofibrosis,
 CC glomerulosclerosis, male pattern baldness, keloids, apocrine cysts, acne,
 CC atherosclerosis, ageing, or a wound (claimed). A method of testing a
 CC candidate composition for PAI-1 inhibition activity comprises
 CC administering the composition to a transgenic animal having a PAI-1 gene
 CC incorporated into its genome, and observing an ameliorating change in the
 CC animal indicative of inhibition of PAI-1 activity, the change being an
 CC improvement of a vascular thrombotic disorder, asthma, chronic
 CC obstructive pulmonary disease, alopecia, undesired weight loss such as
 CC anorexia, Alzheimer's disease, systemic amyloid deposition, systemic
 CC amyloidosis, myelofibrosis, glomerulosclerosis, male pattern baldness,
 CC keloids, apocrine cysts, acne, atherosclerosis, ageing, or a wound (claimed)
 CC hepatosplenomegaly, extramedullary haematopoiesis, or a wound (claimed)
 XX SQ Sequence 2999 BP; 728 A; 810 C; 743 G; 718 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 9; Length 2999;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AACCCCTTAGTCAGATACACTCTTT 25
 Db 2607 AACCCCTTAGTCAGATACACTCTTT 2631
 RESULT 13
 ACF79502
 ID ACF79502 standard; cDNA; 2999 BP.
 XX AC ACF79502;
 XX DT 18-DEC-2003 (first entry)
 XX DE Mouse plasminogen activator inhibitor-1 cDNA.
 XX KW Plasminogen activator inhibitor-1; PAI-1; mouse; transgenic;
 KW thrombolytic; antiasthmatic; antiinflammatory; nootropic;
 KW neuroprotective; antidepressant; nephrotropic; vulnerary;
 KW antisborrheic; dermatological; antiarteriosclerotic; hepatotropic;
 KW gene; ss.
 XX OS Mus musculus.
 XX PH Key Location/Qualifiers
 FT CDS 132..1340
 FT /*tag= a
 FT /product= "Mouse PAI-1"
 XX WO2003071267-A1.
 XX PD 28-AUG-2003.
 XX PF 19-FEB-2003; 2003WO-US005008.
 XX PR 19-FEB-2002; 2002US-0358061P.
 XX PA (UTVA-) UNIV VANDERBILT.
 XX PA (DECL/) DECLERCK P J.
 XX PI Declerck PJ, Vaughan DE, Eren M;
 XX WPI: 2003-721694/68.
 XX P-PSDB; ABR63127.
 XX PT Treating a warm-blooded vertebrate animal having a medical condition,
 PT e.g. Alzheimer's disease, atherosclerosis, myelofibrosis or
 PT glomerulosclerosis, comprises administering a plasminogen activator

PT inhibitor-1 inhibitor.
 XX Disclosure; Page 78-81; 91pp; English.
 XX CC The present sequence is that of cDNA encoding mouse plasminogen activator
 CC inhibitor-1 (PAI-1). The PAI-1 gene can be used to produce transgenic
 CC animals of the invention, useful for screening potential PAI-1
 CC inhibitors. Such PAI-1 inhibitors can be used to treat a warm-blooded
 CC vertebrate animal having a medical condition, e.g. alopecia, undesired
 CC weight loss, Alzheimer's disease, systemic amyloidosis, myelofibrosis,
 CC glomerulosclerosis, male pattern baldness, keloids, apocrine cysts, acne,
 CC atherosclerosis, ageing, or a wound (claimed). A method of testing a
 CC candidate composition for PAI-1 inhibition activity comprises
 CC administering the composition to a transgenic animal having a PAI-1 gene
 CC incorporated into its genome, and observing an ameliorating change in the
 CC animal indicative of inhibition of PAI-1 activity, the change being an
 CC improvement of a vascular thrombotic disorder, asthma, chronic
 CC obstructive pulmonary disease, alopecia, undesired weight loss such as
 CC anorexia, Alzheimer's disease, systemic amyloid deposition, systemic
 CC amyloidosis, myelofibrosis, glomerulosclerosis, male pattern baldness,
 CC keloids, apocrine cysts, acne, atherosclerosis, ageing, or a wound (claimed)
 CC hepatosplenomegaly, extramedullary haematopoiesis, or a wound (claimed)
 XX SQ Sequence 2999 BP; 728 A; 810 C; 743 G; 718 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 9; Length 2999;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AACCCCTTAGTCAGATACACTCTTT 25
 Db 2607 AACCCCTTAGTCAGATACACTCTTT 2631
 RESULT 14
 ABT41815
 ID ABT41815 standard; DNA; 3053 BP.
 XX AC ABT41815;
 XX DT 26-JUN-2003 (first entry)
 XX DE Toxicity modelling related rat gene SEQ ID No 1517.
 XX KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
 KW database; drug screening; toxicity assay; rat; ds.
 XX OS Rattus norvegicus.
 XX PN WO200295000-A2.
 XX PD 28-NOV-2002.
 XX PF 22-MAY-2002; 2002WO-US016173.
 XX PR 22-MAY-2001; 2001US-0292335P.
 PR 13-JUN-2001; 2001US-0297523P.
 PR 19-JUN-2001; 2001US-0298925P.
 PR 10-JUL-2001; 2001US-0303807P.
 PR 10-JUL-2001; 2001US-0303808P.
 PR 10-JUL-2001; 2001US-0303810P.
 PR 28-AUG-2001; 2001US-0315047P.
 PR 27-SEP-2001; 2001US-0324928P.
 PR 22-OCT-2001; 2001US-0330462P.
 PR 01-NOV-2001; 2001US-0330867P.
 PR 21-NOV-2001; 2001US-0331805P.
 PR 06-DEC-2001; 2001US-0336144P.
 PR 19-DEC-2001; 2001US-0340873P.
 PR 21-FEB-2002; 2002US-0357842P.
 PR 21-FEB-2002; 2002US-0357843P.
 PR 21-FEB-2002; 2002US-0357844P.
 PR 15-MAR-2002; 2002US-0364134P.
 PR 08-APR-2002; 2002US-0370144P.

PR 08-APR-2002; 2002US-0370206P.
 PR 08-APR-2002; 2002US-0370247P.
 PR 17-APR-2002; 2002US-0372794P.
 PR 21-APR-2002; 2002US-0371679P.
 XX (GENE-) GENE LOGIC INC.
 PA
 XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 FI WPI; 2003-148464/14.
 XX
 DR Predicting at least one toxic effect of a compound, useful for toxicity
 XX modeling, comprises preparing a gene expression profile of a tissue or
 PT cell sample exposed to the compound, and comparing the gene expression
 PT profile to a database.
 PT
 XX
 PS Example 4; Page; 446pp; English.
 XX
 CC The invention relates to a novel method of predicting at least one toxic
 CC effect of a compound. The method comprises a gene expression profile of a
 CC tissue or cell sample exposed to the compound, and comparing the gene
 CC expression profile to a database comprising at least part of the data or
 CC information given in the specification. The methods are useful for
 CC predicting at least one toxic effect of a compound, predicting the
 CC progression of a toxic effect of a compound, predicting the renal
 CC toxicity of a compound, or identifying toxicity markers in tissues or
 CC cells exposed to known renal toxin. The genes are useful as toxicity
 CC markers in drug screening and toxicity assays, in monitoring disease or
 CC physiological states, or disease progression. This polynucleotide
 CC represents a rat DNA sequence relating to the toxic effect database
 CC described in the specification. NOTE: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the World Intellectual Property
 CC Organization
 XX
 SQ Sequence 3053 BP; 773 A; 806 C; 762 G; 712 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 7; Length 3053;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AACCCCTTAGTCAGATACACTCTTT 25
 DB 2665 AACCCCTTAGTCAGATACACTCTTT 2689
 RESULT 15
 ACF79503
 ID ACF79503 standard; cDNA; 3053 BP.
 XX
 AC ACF79503;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Rat plasminogen activator inhibitor-1 cDNA.
 XX
 KW Plasminogen activator inhibitor-1; PAI-1; rat; transgenic; thrombolytic;
 KW antiasthmatic; antiinflammatory; nootropic; neuroprotective;
 KW antidepressant; nephrotropic; vulnary; antiseborrheic; dermatological;
 KW antiarteriosclerotic; hepatotropic; gene; ss.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 CDS 119..1327
 FT /*tag= a
 FT /product= "Rat PAI-1"
 XX
 PN WO2003071267-A1.
 XX
 PD 28-AUG-2003.
 XX
 PF 19-FEB-2003; 2003WO-US005008.

XX 19-FEB-2002; 2002US-0358061P.
 XX (UYVA-) UNIV VANDERBILT.
 PA (DECL/) DECLERCK P J.
 XX
 FI Declerck PJ, Vaughan DE, Eren M;
 XX WPI; 2003-721694/68.
 DR
 XX
 PT Treating a warm-blooded vertebrate animal having a medical condition,
 PT e.g. Alzheimer's disease, atherosclerosis, myelofibrosis or
 PT glomerulosclerosis, comprises administering a plasminogen activator
 PT inhibitor-1 inhibitor.
 XX
 PS Disclosure; Page 84-86; 9lpp; English.
 XX
 CC The present sequence is that of cDNA encoding rat plasminogen activator
 CC inhibitor-1 (PAI-1). The PAI-1 gene can be used to produce transgenic
 CC animals of the invention, useful for screening potential PAI-1
 CC inhibitors. Such PAI-1 inhibitors can be used to treat a warm-blooded
 CC vertebrate animal having a medical condition, e.g. alopecia, undesired
 CC weight loss, Alzheimer's disease, systemic amyloidosis, myelofibrosis,
 CC glomerulosclerosis, male pattern baldness, keloids, apocrine cysts, acne,
 CC atherosclerosis, ageing, or a wound (claimed). A method of testing a
 CC candidate composition for PAI-1 inhibition activity comprises
 CC administering the composition to a transgenic animal having a PAI-1 gene
 CC incorporated into its genome, and observing an ameliorating change in the
 CC animal indicative of inhibition of PAI-1 activity, the change being an
 CC improvement of a vascular thrombotic disorder, asthma, chronic
 CC obstructive pulmonary disease, alopecia, undesired weight loss such as
 CC anorexia, Alzheimer's disease, systemic amyloid deposition, systemic
 CC amyloidosis, myelofibrosis, glomerulosclerosis, male pattern baldness,
 CC keloids, apocrine cysts, acne, atherosclerosis, ageing,
 CC hepatosplenomegaly, extramedullary haematopoiesis, or a wound (claimed)
 XX
 SQ Sequence 3053 BP; 773 A; 806 C; 762 G; 712 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 9; Length 3053;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AACCCCTTAGTCAGATACACTCTTT 25
 DB 2665 AACCCCTTAGTCAGATACACTCTTT 2689
 Search completed: March 24, 2004, 22:28:23
 Job time : 134.899 secs

ECORI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.
 Vector: pCS107; Site 1: EcoRI; Site 2: NotI
 Host: Escherichia coli XL1-blue.

FEATURES

source

1..870
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TEgg059102"
 /dev_stage="egg"
 /lab_host="Escherichia coli XL1-blue"
 /clone_lib="XGC-egg"
 /note="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from sug of poly A+ RNA from egg.
 EcorI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 80.8%; Score 20.2; DB 13; Length 870;
 Best Local Similarity 88.0%; Pred. No. 1.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGATGACTACTTT 25
 Db 283 ACCAACCTTAGTCAGATACTAATTT 307

RESULT 2

LOCUS

DEFINITION A0586782 467 bp DNA linear GSS 07-JUN-1999
 RPCI-11-433B6.TV RPCI-11 Homo sapiens genomic clone RPCI-11-433B6,
 genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
 Map Building
 Unpublished (1997)
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbs@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genet cs (info@resgen.com). BAC end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: 77

Class: BAC ends.

Location/Qualifiers

1..467
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GDB:7665989"
 /db_xref="taxon:9606"
 /clone="RPCI-11-433B6"
 /sex="Male"
 /cell_type="Lymphocytes"
 /clone_lib="RPCI-11"
 /note="vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPCI11 Human Male BAC Library"

FEATURES

source

Query Match 77.6%; Score 19.4; DB 9; Length 549;
 Best Local Similarity 95.2%; Pred. No. 3.3e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ACCACCTTAGTCAGATACTAC 22
 Db 461 ACCACCTTAGTCAGATACTAC 441

RESULT 4

LOCUS

DEFINITION A1349205 244 bp mRNA linear EST 16-FEB-1999
 ta74b03.x2 NCI_CGAP_HSC2 Homo sapiens cDNA clone IMAGE:2049773 3',
 mRNA sequence.

ACCESSION

ORIGIN

Query Match 79.2%; Score 19.8; DB 28; Length 467;
 Best Local Similarity 91.3%; Pred. No. 2.1e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CCACCTTAGTCAGATACTACTTT 25
 Db 428 CCACCTTAGTCAGATACTACTTT 450

RESULT 3

LOCUS

DEFINITION A1188862 549 bp mRNA linear EST 28-OCT-1998
 qd27a12.x1 Soares placenta 8to9weeks 2NbHP8to9W Homo sapiens cDNA
 clone IMAGE:1724926 3' similar to gb:X15414 ALDOSE REDUCTASE
 (HUMAN); mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strauberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Trace considered overall poor quality
 Insert Length: 1099 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1..549
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1724926"
 /dev_stage="two placentae: one from 8 weeks and another
 from 9 weeks post conception"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares placenta 8to9weeks 2NbHP8to9W"
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGAGCGCGCGATTTTTTTTTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M.Fatima Bonaldo."

FEATURES

source

Query Match 77.6%; Score 19.4; DB 9; Length 549;
 Best Local Similarity 95.2%; Pred. No. 3.3e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

Qy 2 ACCACCTTAGTCAGATACTAC 22
 Db 461 ACCACCTTAGTCAGATACTAC 441

RESULT 4

LOCUS

DEFINITION A1349205 244 bp mRNA linear EST 16-FEB-1999
 ta74b03.x2 NCI_CGAP_HSC2 Homo sapiens cDNA clone IMAGE:2049773 3',
 mRNA sequence.

ACCESSION

```

VERSION      AI349205.1  GI:4086411
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 244)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
             National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,
             M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 347 Std Error: 0.00
             Seq primer: -40UP from Gibco.
FEATURES     Location/Qualifiers
             1..244
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:2049773"
             /tissue_type="stem cell 34+/38+"
             /dev_stage="adult"
             /lab_host="DH10B"
             /clone_lib="NCI CGAP HSC2"
             /note="Organ: bone marrow; Vector: pAMP1; mRNA made from
             bone marrow; stem cells 34+/38+, cDNA made by oligo-dT
             priming. Directionally cloned. Size-selected on agarose
             gel, average insert size 400 bp. Primary library,
             non-amplified."
ORIGIN
Query Match 76.8%; Score 19.2; DB 9; Length 244;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCACCTTAGTCAGATACACTTT 25
||||| ||||||| |||||
Db 104 ACCACCTTGTGCAGATACACTTT 127

RESULT 5
CF793467 569 bp mRNA linear EST 21-OCT-2003
LOCUS
DEFINITION 888857 MARC 4PTG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION CF793467
VERSION
KEYWORDS
SOURCE
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 569)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Noneman,D.J., Wray,J.E. and Keele,J.W.
TITLE Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and

```

```

trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMW8011 row: M column: 8
Seq primer: GTAATACGACTCACTATAGG.
FEATURES     Location/Qualifiers
             1..569
             /organism="Sus scrofa"
             /mol_type="mRNA"
             /db_xref="taxon:9823"
             /tissue_type="pooled"
             /lab_host="DH10B"
             /clone_lib="MARC 4PTG"
             /note="Vector: pCDNA3.1; Site_1: EcoRI; Site_2: NotI;
             library made with combined RNA from day-10, day-13,
             day-15, day-25, and day-30 whole embryos."
ORIGIN
Query Match 75.2%; Score 18.8; DB 14; Length 569;
Best Local Similarity 90.9%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTAC 22
||||| ||||||| |||||
Db 33 AACCACTTAGTCAGATACACTAC 54

RESULT 6
H85044/c
LOCUS
DEFINITION H85044.1 GI:1063699
ACCESSION H85044
VERSION
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 319)
AUTHORS Hillier,L., Clark,N., Dubucq,T., Eliiston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisan,S., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 299
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: Promega -21m13
High quality sequence stop: 299.
FEATURES     Location/Qualifiers
             1..319
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="GDB:384772"
             /db_xref="taxon:9606"
             /clone="IMAGE:219731"
             /sex="male"
             /tissue_type="retina"
             /dev_stage="55 year old"
             /lab_host="DH10B (ampicillin resistant)"
             /clone_lib="Soares retina N2b4HR"
             /note="Organ: eye; Vector: pT7n3D (Pharmacia) with a
             modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

```

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p7T3 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)⁺ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Patima Bonaldo. "

ORIGIN

Query Match 74.4%; Score 18.6; DB 14; Length 319;
Best Local Similarity 84.0%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTACTGACATACACTTTT 25

Db 139 AGCCACCTTAACATGACACACTTT 115

RESULT 7
BY171910 364 bp mRNA linear EST 10-DEC-2002
LOCUS
DEFINITION
BY171910 RIKEN full-length enriched, bone marrow macrophage Mus
musculus cDNA clone I83090J19 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 364)
Osato, N., Furuno, M., Kaakawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Kiyosawa, H., Gojoubori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayase, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Omo, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry
and Microbiology/Parasitology Institute for Molecular Bioscience
University of Queensland Brisbane, Q 4072 Australia) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
source

Location/Qualifiers
1..364
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I83090J19"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 364;
Best Local Similarity 84.0%; Pred. No. 6.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTACTGACATACACTTTT 25

Db 28 AACCACTTACTGACATACACTTTT 52

RESULT 8

BY171910

LOCUS

DEFINITION

BY171910 RIKEN full-length enriched, bone marrow macrophage Mus

musculus cDNA clone I83090J19 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 377)

Osato, N., Furuno, M., Kaakawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Kiyosawa, H., Gojoubori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayase, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry
and Microbiology/Parasitology Institute for Molecular Bioscience
University of Queensland Brisbane, Q 4072 Australia) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

Location/Qualifiers
1..364
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I83090J19"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"

Query Match 74.4%; Score 18.6; DB 13; Length 364;
Best Local Similarity 84.0%; Pred. No. 6.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTACTGACATACACTTTT 25

Db 28 AACCACTTACTGACATACACTTTT 52

RESULT 8

BY171910

LOCUS

DEFINITION

BY171910 RIKEN full-length enriched, bone marrow macrophage Mus

musculus cDNA clone I83090J19 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 377)

Osato, N., Furuno, M., Kaakawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Kiyosawa, H., Gojoubori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayase, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@genome.gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by David A. Hume (Depts. of Biochemistry
 and Microbiology/Parasitology Institute for Molecular Bioscience
 University of Queensland Brisbane, Q 4072 Australia) whose
 assistance we gratefully acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source

```

1. .406
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I830153E12"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"

```

ORIGIN

```

Query Match      74.4%; Score 18.6; DB 13; Length 406;
Best Local Similarity 84.0%; Pred. No. 6.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 AACCACTTAGTCAGATACACTTTT 25
    |||||
Db 12 AACCACTTAGTCAGATACACTTTT 36

```

RESULT 11

```

BY688349 LOCUS 411 bp mRNA linear EST 16-DEC-2002
DEFINITION BY688349 RIKEN full-length enriched, stroma cell Mus musculus cDNA
clone I320028N18 3', mRNA sequence.
ACCESSION BY688349
VERSION BY688349.1 GI:27093998
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

```

ORGANISM

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 411)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,

```

FEATURES

```

source
1. .411
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I320028N18"
/cell_type="stroma cell"
/clone_lib="RIKEN full-length enriched, stroma cell"

```

ORIGIN

Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,
 Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, I.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, K.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@genome.gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Takashi Ishikawa (Department of Surgery
 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama
 236-0004 Japan) whose assistance we gratefully acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

```

Query Match      74.4%; Score 18.6; DB 13; Length 411;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTCTTT 25
    |||||
DB 16 AACCACTTAGTCAGATACACTCTTT 40

RESULT 12
BY573260
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 413)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawajiri, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
DOI: 10.1038/420563a
12466851

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to

```

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

```

Location/Qualifiers
1..413
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="taxon:10090"
/clone="F830301I21"
/tissue_type="activated spleen"
/clone_lib="RIKEN full-length enriched, activated spleen"

```

ORIGIN

```

Query Match      74.4%; Score 18.6; DB 13; Length 413;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTCTTT 25
    |||||
DB 27 AACCACTTAGTCAGATACACTCTTT 51

```

RESULT 13

```

BY513444
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

BY513444 RIKEN full-length enriched, bone marrow macrophage Mus
musculus cDNA clone 1830149K22 3', mRNA sequence.

```

```

BY513444.1 GI:26847823
EST.
Mus musculus (house mouse)
Mus musculus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 414)

```

```

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawajiri, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,

```

Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

Location/Qualifiers

1..414

source

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="I830148K22"

/tissue_type="bone marrow"

/cell_type="macrophage"

/clone_lib="RIKEN full-length enriched, bone marrow macrophage"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 414;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGACTACTTCTT 25

|||||
Db 20 AACCACTTACTGACTACTTCTT 44

RESULT 14

BY503440

LOCUS

BY503440 417 bp mRNA linear EST 14-DEC-2002

DEFINITION

BY503440 RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone I830081H09 3', mRNA sequence.

ACCESSION

BY503440

VERSION

1 GI:26837819

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (Bases 1 to 417)

AUTHORS

Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 417)
Okazaki, Y., Furuno, M., Kaekawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Shoyabach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusica, V., Ciothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Partea, G., Resole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynehaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:53:36 ; Search time 906.607 Seconds
(without alignments)
1960.127 Million cell updates/sec

Title: US-09-889-611a-60_COPY_2_42

Perfect score: 41

Sequence: 1 gaagaaactaatacaacc.....ccttagtcagatactacttt 41

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sv.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39.4	96.1	4229	9	AF234618	AF234618 Homo sapi
2	39.4	96.1	157284	9	AC072051	AC072051 Homo sapi
3	39.4	96.1	189092	2	AC027524	AC027524 Homo sapi
4	33.6	82.0	6182	6	AX346917	AX346917 Sequence
5	29.4	71.7	207510	2	AC147488	AC147488 Otlemur
6	28.2	68.8	198669	10	AC103453	AC103453 Rattus no
7	28.2	68.8	219071	2	AC133259	AC133259 Rattus no
8	25	61.0	6182	6	AX346916	AX346916 Sequence
9	25	61.0	121538	10	AC129295	AC129295 Mus muscu
10	25	61.0	200844	2	AC134414	AC134414 Mus muscu
11	25	61.0	213457	2	AC110569	AC110569 Mus muscu
12	24.6	60.0	160615	10	AL672023	AL672023 Mouse DNA
13	24	58.5	3473	9	HS0804835	AL833522 Homo sapi
14	24	58.5	5276	6	AX251133	AX251133 Sequence
15	24	58.5	5276	6	AX345728	AX345728 Sequence
16	24	58.5	146746	9	AL353718	AL353718 Human DNA
17	24	58.5	222097	2	AC129874	AC129874 Rattus no
18	24	58.5	228783	2	AC094377	AC094377 Rattus no
19	24	58.5	235965	2	AC127103	AC127103 Rattus no
20	23.6	57.6	148648	9	AC021713	AC021713 Homo sapi
21	23.4	57.1	99539	8	AC113332	AC113332 Oryza sat
22	23.4	57.1	122599	8	AC130605	AC130605 Oryza sat
23	23.4	57.1	161161	5	AL935146	AL935146 Zebrafish
24	23.4	57.1	165700	10	AC132342	AC132342 Mus muscu
25	23.4	57.1	233830	2	AC141892	AC141892 Mus muscu
26	23.4	57.1	289080	2	AC111582	AC111582 Rattus no
27	23.2	56.6	1005	3	AY005308	AY005308 Xylocopa
28	23.2	56.6	132093	2	AC141964	AC141964 Rattus no
29	23.2	56.6	260121	2	AC130738	AC130738 Rattus no
30	23.2	56.6	338614	2	AC103268	AC103268 Rattus no
31	23	56.1	4038	9	HS24P521	X67640 H.sapiens H
32	23	56.1	69037	10	AC003949	AC003949 Mus muscu
33	23	56.1	116103	9	AL590653	AL590653 Human DNA
34	23	56.1	123851	9	AC113410	AC113410 Homo sapi
35	23	56.1	131886	2	AC020834	AC020834 Mus muscu
36	23	56.1	158296	2	BX530098	BX530098 Danio rer
37	23	56.1	198694	10	BX072557	BX072557 Mouse DNA
38	23	56.1	216921	2	BX322658	BX322658 Mus muscu
39	23	56.1	235952	2	BX784023	BX784023 Danio rer
40	23	56.1	346601	2	AC111981	AC111981 Rattus no
41	22.8	55.6	168169	2	AC145953	AC145953 Pan trogl
42	22.8	55.6	196870	2	AC115209	AC115209 Rattus no
43	22.8	55.6	201286	2	AC102228	AC102228 Mus muscu
44	22.8	55.6	219817	2	AC099196	AC099196 Rattus no
45	22.8	55.6	220129	2	AC096823	AC096823 Rattus no

ALIGNMENTS

RESULT 1	AF234618	4229 bp	DNA	linear	PRI 29-MAR-2002
LOCUS	AF234618	Homo sapiens megain gene, promoter region and partial sequence.			
DEFINITION	AF234618				
ACCESSION	AF234618.2	GI:19808130			
VERSION					
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 4229)				
AUTHORS	Miyata,T., Nangaku,M., Inagi,R. and Kurokawa,K.				
TITLE	Transcriptional regulation of a mesangium-predominant gene, megain				
JOURNAL	Unpublished				

```

REFERENCE
AUTHORS      Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
TITLE        Direct Submission
JOURNAL      Submitted (15-FEB-2000) Institute of Medical Sciences and
              Department of Internal Medicine, Tokai University School of
              Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
              (bases 1 to 4229)
REFERENCE
AUTHORS      Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
TITLE        Direct Submission
JOURNAL      Submitted (29-MAR-2002) Institute of Medical Sciences and
              Department of Internal Medicine, Tokai University School of
              Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
REMARK       Sequence update by submitter
COMMENT      On Mar 29, 2002 this sequence version replaced gi:18000453.
FEATURES
source       Location/Qualifiers
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="18"
             /map="18q21.3"
             1..4021
             4022..>4229
             /product="megsin"
ORIGIN
Query Match      96.1%; Score 39.4; DB 9; Length 4229;
Best Local Similarity 97.6%; Pred. No. 7.4e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 GAATGAACATCAATCAACACCACTTACTGACACTACTTT 41
Db      3920 GAATGAACATCAATCAACACCACTTACTGACACTACTTT 3960
RESULT 2
AC072051/c      157284 bp DNA linear PRI 03-DEC-2001
LOCUS           Homo sapiens chromosome, clone RP11-79D21, complete sequence.
DEFINITION      AC072051
ACCESSION       AC072051
VERSION         AC072051.8 GI:16974280
KEYWORDS        HTG.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS        Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE          Homo sapiens chromosome, clone RP11-79D21
JOURNAL        Unpublished
REFERENCE
AUTHORS        Birren, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N.,
              Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
              Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
              Campopiano, A., Castelle, A., Choepel, Y., Colangelo, M., Collins, S.,
              Collumore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
              Dodge, S., Domingo, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
              Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
              Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
              Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
              Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
              Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
              McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
              Meldrim, J., Meneus, J., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
              Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
              O'Neil, D., Oliver, J., Norman, C.H., O'Connor, T., Peterson, K.,
              Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
              Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
              Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
              Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
              Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
              Young, G., Zainoun, J., Zimmer, A. and Zody, M.
              Direct Submission
FEATURES
source

```

JOURNAL
REFERENCE
AUTHORS

Submitted (07-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 157284)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fargo, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
 Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 157284)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fargo, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (03-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All 18 repeats were identified using RepeatMasker:
 Smit, A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seg.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L966
 Center clone name: 79_D_21
 ----- Location/Qualifiers
 1..157284
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Rp11-79D21"

FEATURES
source

```
repeat_region /clone lib="RPCI-11 Human Male BAC"
repeat_region 599..635
repeat_region /rpt_family="(T)n"
repeat_region 1040..1077
repeat_region /rpt_family="AT_rich"
repeat_region 1411..1747
repeat_region /rpt_family="L1P"
repeat_region 2632..2992
repeat_region /rpt_family="MLT1A1"
repeat_region 3115..3138
repeat_region /rpt_family="AT_rich"
repeat_region complement(3487..3828)
repeat_region /rpt_family="L1MC4a"
repeat_region 3834..4338
repeat_region /rpt_family="MLT1C"
repeat_region complement(4339..5397)
repeat_region /rpt_family="L1MC4a"
repeat_region complement(5742..5899)
repeat_region /rpt_family="MER5B"
repeat_region 6499..6816
repeat_region /rpt_family="HAL1"
repeat_region complement(7416..7621)
repeat_region /rpt_family="MER3"
repeat_region complement(8163..9090)
repeat_region /rpt_family="L2"
repeat_region 9276..9419
repeat_region /rpt_family="MIR"
repeat_region 9598..9658
repeat_region /rpt_family="MER81"
repeat_region 9659..10033
repeat_region /rpt_family="MLT1J"
repeat_region 13986..14009
repeat_region /rpt_family="AT_rich"
repeat_region 14011..14115
repeat_region /rpt_family="(TA)n"
repeat_region complement(14942..15044)
repeat_region /rpt_family="MIR3"
repeat_region 15156..15298
repeat_region /rpt_family="L2"
repeat_region 15592..15773
repeat_region /rpt_family="L1M4c"
repeat_region 15874..18310
repeat_region /rpt_family="L1PA16"
repeat_region 16373..16374
unsure /note="1328 bp of bacterial transposon insertion in BAC
excised between these 2 bp"
repeat_region 18311..18331
repeat_region /rpt_family="(TAAA)n"
repeat_region 18332..19566
repeat_region /rpt_family="L1PA16"
repeat_region 19567..19587
repeat_region /rpt_family="(CA)n"
repeat_region 19588..20042
repeat_region /rpt_family="L1PA16"
repeat_region 20043..20064
repeat_region /rpt_family="AT_rich"
repeat_region complement(21259..21386)
repeat_region /rpt_family="MIR3"
repeat_region 22395..22420
repeat_region /rpt_family="AT_rich"
repeat_region complement(22788..22948)
repeat_region /rpt_family="FRAM"
repeat_region complement(23843..24253)
repeat_region /rpt_family="MSTA"
repeat_region complement(24678..25041)
repeat_region /rpt_family="MLT1A2"
repeat_region 25054..25094
repeat_region /rpt_family="HAL1"
repeat_region 25214..25358
repeat_region /rpt_family="L1M4c"
repeat_region complement(25602..25854)
repeat_region /rpt_family="MIR"
repeat_region 26746..26796
```

```
repeat_region /rpt_family="AT_rich"
repeat_region 27451..27484
repeat_region complement(28183..28219)
repeat_region /rpt_family="MIR"
repeat_region complement(28249..28325)
repeat_region /rpt_family="MIR"
repeat_region complement(29372..29459)
repeat_region /rpt_family="MIR"
repeat_region 29539..30003
repeat_region /rpt_family="MER6B"
repeat_region 30596..30624
repeat_region /rpt_family="AT_rich"
repeat_region 31418..31829
repeat_region /rpt_family="L2"
repeat_region 31846..31915
repeat_region /rpt_family="(TGA)n"
repeat_region complement(32019..32103)
```

Query Match 96.1%; Score 39.4; DB 9; Length 157284;
Best Local Similarity 97.6%; Pred. No. 5.4e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACCTTAGTCAGACTACTACTTT 41
|||||
Db 155067 GAATGAACATACATACACACCTTAGTCAGACTACTTT 155027

RESULT 3
AC027524
LOCUS Homo sapiens chromosome 18 clone RP11-317G1 map 18, 3 unordered
DEFINITION AC027524 189092 bp DNA linear HTG 27-MAR-2003
ACCESSION AC027524.4 GI:21307437
VERSION HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-317G1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 189092)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.903019
Consensus quality: 207213 bases at least Q40
Consensus quality: 207263 bases at least Q30
Consensus quality: 207289 bases at least Q20

Insert size: 218000; agarose-gel
 Insert size: 207410; sum-of-contigs
 Quality coverage: 14.36x in Q20 bases; agarose-gel
 Quality coverage: 15.10x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 38181: contig of 38181 bp in length
 * 38182 38281: gap of unknown length
 * 38282 207510: contig of 169229 bp in length.

FEATURES

source
 Location/Qualifiers
 1..207510
 /organism="Otolemur garnettii"
 /mol_type="genomic DNA"
 /db_xref="taxon:30611"
 /clone="CH256-273K16"
 /clone_lib="CH256"
 1..38181
 /notes="assembly_fragment"
 clone_end:T7
 vector_side:left
 38282..207510
 /notes="assembly_fragment"
 clone_end:SP6
 vector_side:right"

misc_feature
 1..38181
 /notes="assembly_fragment"
 clone_end:T7
 vector_side:left
 38282..207510
 /notes="assembly_fragment"
 clone_end:SP6
 vector_side:right"

misc_feature
 38282..207510
 /notes="assembly_fragment"
 clone_end:SP6
 vector_side:right"

ORIGIN

Query Match 71.7%; Score 29.4; DB 2; Length 207510;
 Best Local Similarity 84.6%; Pred. No. 0.44;
 Matches 33; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACCTTGTAGTCAGACTACT 39
 |||||
 Db 151706 GAATGAACCATATGATGCTGCTTGTAGTCAGACTACT 151668

RESULT 6
 AC103453 199669 bp DNA linear ROD 07-JUN-2003
 LOCUS Rattus norvegicus 13 BAC CH230-127N19 (Children's Hospital Oakland
 DEFINITION Research Institute) complete sequence.
 AC103453
 VERSION HTG.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS Muzny,D,Marle., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Hayes,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Garcia,R., Garcia,A., Garner,T., Garza,M.,
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,U.,
 Lorenschew,L., Louisege,H., Lozano,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawhinney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
 Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
 Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
 Newton,N., Nguyen,N., Norris,S., Nwaokeme,O., Okwuonu,G.,
 Olarunpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
 Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A.,
 Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
 Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
 Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
 Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
 Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
 Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
 Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
 Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
 Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
 Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
 Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlezyk,R.,
 Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,X.,
 Yen,J., Yoon,L., Yoon,F., Zhang,J., Zhou,J., Zhou,X.,
 Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
 Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 199669)
 Worley,K.C.
 Direct Submission
 Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 199669)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 199669)
 Worley,K.C.
 Direct Submission
 Submitted (07-JUN-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 7, 2003 this sequence version replaced gi:30521373.
 Sequencing is completed to a minimum standard of double strand
 coverage with a minimum of 2 clones and 2 reads with no ambiguities
 or 2 chemistries with a minimum of 2 clones and 3 reads with no
 ambiguities. If the sequence quality does not meet this standard,
 it will be indicated in the annotation.

Location/Qualifiers
 1..199669
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /chromosome="13"
 /clone="CH230-127N19"
 5394..5759
 /rpt_family="Lx2_3"
 5760..5861
 /rpt_family="(CA)n"
 5862..6447
 /rpt_family="Lx2_3"
 6522..6586
 /rpt_family="(TG)n"
 6611..6938

repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region

FEATURES
 source
 Location/Qualifiers
 1..199669
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /chromosome="13"
 /clone="CH230-127N19"
 5394..5759
 /rpt_family="Lx2_3"
 5760..5861
 /rpt_family="(CA)n"
 5862..6447
 /rpt_family="Lx2_3"
 6522..6586
 /rpt_family="(TG)n"
 6611..6938

Lorensuhea, L., Loulsegad, H., Lozado, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindarne, M., Mamoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G.S., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, J., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackelme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wiczzyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 219071)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (09-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 219071)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:22759157.

The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KGN7

Center clone name: CH230-318J10

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 194595 bases at least Q40

Consensus quality: 198491 bases at least Q30

Consensus quality: 200795 bases at least Q20

Estimated insert size: 202052; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 217199: contig of 217199 bp in length

* 217200 217299: gap of unknown length

* 217300 219071: contig of 1772 bp in length.

FEATURES

source

1..219071

/organism="Rattus norvegicus"

/mol_type="Genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-318J10"

1..1340

/note="wgs_contig"

3673..5446

/note="wgs_contig"

152568..153448

/note="clone_boundary"

clone_end:77

site:

end sequence:BZ142608"

212439..213833

/note="wgs_end_extension"

clone_end:77"

ORIGIN

Query Match 68.8%; Score 28.2; DB 2; Length 219071;

Best Local Similarity 80.5%; Pred. No. 1.3;

Matches 33; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATAACACACCTTAGTCAGATACACTTTT 41

|||||
 DB 198951 GAATGACACATACATACATCTTTAATCATATGCTATTTT 198911

RESULT 8

AX346916

LOCUS

Sequence 1987 from Patent WO0200928.

DEFINITION

ACCESSION

VERSION

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Epigenomics AG (DE)

Location/Qualifiers

1..6182

/organism="synthetic construct"

/mol_type="synthetic DNA"

/db_xref="taxon:32630"

/note="chemically treated genomic DNA (Homo sapiens)"

FEATURES

source

1..6182

/organism="synthetic construct"

/mol_type="synthetic DNA"

/db_xref="taxon:32630"

/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match

Best Local Similarity

Matches

QY 1 GAATGAACCTACATAACACACCTTAGTCAGATACACTTTT 41

|||||
 DB 4872 GAATGAATATATATATATTTTAGTAGATATATTTT 4912

RESULT 9

AC129295

LOCUS

Mus musculus BAC clone RP24-560M23 from chromosome 1, complete

DEFINITION

AC129295

121538 bp DNA linear

ROD 27-NOV-2003

```

sequence.
AC129295
VERSION AC129295.4 GI:37651860
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Wang, C., Haakerson, W. and Haglund, K.
The sequence of Mus musculus BAC clone RP24-560M23
Unpublished (2001)
REFERENCE 2 (bases 1 to 121538)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 121538)
AUTHORS McPherson, J. D. and Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 121538)
AUTHORS McPherson, J. D. and Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 121538)
AUTHORS Wilson, R. K.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 121538)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 11, 2003 this sequence version replaced gi:28867176.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0560M23
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

```

```

FEATURES
    source
        Location/Qualifiers
            1..121538
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /chromosome="1"
                /map="1"
                /clone="RP24-560M23"
                /clone_lib="RPCI-24"
            1..1753
                /rpt_family="L1"
            1801..1935
                /rpt_family="L1"
            5774..6000
                /rpt_family="L1"
            6013..6221
                /rpt_family="MaLR"
            6213..6690
                /rpt_family="MaLR"
            6830..7037
                /rpt_family="L1"
            7108..7293
                /rpt_family="B2"
            7347..7481
                /rpt_family="L1"
            7501..7597
                /rpt_family="MaLR"
            7663..8150
                /rpt_family="L1"
            8914..9125
                /rpt_family="MaLR"
            9341..9526
                /rpt_family="L1"
            10108..10198
                /rpt_family="L1"
            10274..10520
                /rpt_family="B4"
            10643..10724
                /rpt_family="L1"
            10974..11136
                /rpt_family="L1"
            11519..11625
                /rpt_family="ERVK"
            12864..13008
                /rpt_family="RMER15"
            13045..13270
                /rpt_family="RMER15"
            13801..13978
                /rpt_family="L1"
            14537..14672
                /rpt_family="L1"
            14909..15475
                /rpt_family="ERV1"
            15664..16263
                /rpt_family="L1"
            16711..16946
                /rpt_family="ERVK"
            17455..18017
                /rpt_family="L1"
            18013..18607
                /rpt_family="L1"
            18665..18990
                /rpt_family="ERVK"
            19007..19123
                /rpt_family="L1"
            19578..19724
                /rpt_family="B2"
            19909..19963
                /rpt_family="MaLR"
            19961..20188
                /rpt_family="MaLR"
            20307..20846
                /rpt_family="MaLR"

```

```

repeat_region 21301..22200
/rpt_family="L1"
repeat_region 29157..29428
/rpt_family="L1"
repeat_region 29859..31024
/rpt_family="L1"
repeat_region 31028..31144
/rpt_family="L1"
repeat_region 31170..32186
/rpt_family="L1"
repeat_region 32259..32415
/rpt_family="L1"
repeat_region 32867..33012
/rpt_family="Alu"
repeat_region 33079..33191
/rpt_family="L1"
repeat_region 33249..34124
/rpt_family="L1"
repeat_region 34600..35070
/rpt_family="L1"
repeat_region 35469..35546
/rpt_family="L1"
repeat_region 35547..35886
/rpt_family="Mair"
repeat_region 35887..36282
/rpt_family="L1"
repeat_region 36487..36745
/rpt_family="L1"
repeat_region 37707..38105
/rpt_family="Mair"
repeat_region 38106..38239
/rpt_family="B4"
repeat_region 38398..38463
/rpt_family="L2"
repeat_region 39337..39455
/rpt_family="Alu"
repeat_region 39511..39757
/rpt_family="MER1_type"
repeat_region 40057..40451
/rpt_family="Mair"
repeat_region 40638..41045
/rpt_family="RMER17C"
repeat_region 43612..43654
/rpt_family="ERV1"
repeat_region 43676..43914
/rpt_family="ERVK"
repeat_region 44000..44432
/rpt_family="RMER17C"
repeat_region 45173..45288
/rpt_family="L1"
repeat_region 45732..46123
/rpt_family="Mair"

Query Match 51.0%; Score 25; DB 10; Length 121538;
Best Local Similarity 75.8%; Pred. No. 25;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACACACCTTAGTCAGATACACTTT 41
Db 115259 GAATGAACATACATACATACATCTTTATACAGGCTATTT 115299

RESULT 10
AC134414/C
LOCUS AC134414
DEFINITION Mus musculus clone RP24-70H12, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC134414
VERSION AC134414.3 GI:28565730
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 208644)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gird,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208644)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Faro,S.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Gardyna,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramaamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 26, 2003 this sequence version replaced gi:28275018.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26735
Center clone name: 70_H12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197943 bases at least Q40
Consensus quality: 199015 bases at least Q30
Consensus quality: 199440 bases at least Q20
Insert size: 194000; agarose-ep
Insert size: 199644; sum-of-contigs
Quality coverage: 9.2 in Q20 bases; agarose-ep

```

Quality coverage: 8.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 54386: contig of 54386 bp in length
* 54387 5486: gap of 100 bp
* 5487 55846: contig of 1360 bp in length
* 55847 55946: gap of 100 bp
* 55947 59197: contig of 3251 bp in length
* 59198 59297: gap of 100 bp
* 59298 64542: contig of 5245 bp in length
* 64543 64642: gap of 100 bp
* 64643 69113: contig of 4471 bp in length
* 69114 69213: gap of 100 bp
* 69214 72896: contig of 3683 bp in length
* 72897 72996: gap of 100 bp
* 72997 78283: contig of 5287 bp in length
* 78284 78384: gap of 100 bp
* 78384 90557: contig of 12174 bp in length
* 90558 90658: gap of 100 bp
* 90658 10316: contig of 10659 bp in length
* 10317 10416: gap of 100 bp
* 10417 120870: contig of 19454 bp in length
* 120871 120970: gap of 100 bp
* 120971 145380: contig of 24410 bp in length
* 145381 145480: gap of 100 bp
* 145481 174041: contig of 28561 bp in length
* 174042 174141: gap of 100 bp
* 174142 200844: contig of 26703 bp in length.

```

FEATURES

```

source
1. 200844
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /clone="RP24-70H12"
   /clone_lib="RPCI-24 Male Mouse BAC"
misc_feature
1. 54386
   /note="assembly_fragment"
   /note="end:SP6"
   vector_side:left
misc_feature
54487..55846
   /note="assembly_fragment"
misc_feature
55947..59197
   /note="assembly_fragment"
misc_feature
59298..64542
   /note="assembly_fragment"
misc_feature
64643..69113
   /note="assembly_fragment"
misc_feature
69214..72896
   /note="assembly_fragment"
misc_feature
72997..78283
   /note="assembly_fragment"
misc_feature
78384..90557
   /note="assembly_fragment"
misc_feature
90658..101316
   /note="assembly_fragment"
misc_feature
101417..120870
   /note="assembly_fragment"
misc_feature
120971..145380
   /note="assembly_fragment"
misc_feature
145481..174041
   /note="assembly_fragment"
misc_feature
174142..200844
   /note="assembly_fragment"
   /note="end:T7"
   vector_side:right

```

ORIGIN

```

Query Match      61.0%; Score 25; DB 2; Length 200844;
Best Local Similarity 75.6%; Pred. No. 24;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 1 GAATGACTACATTAACCAACACCTTACTGACACTACTCTT 41
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60430 GAATACCAACAGAAATTAAGACCTTGCTCAATATTACTTT 60390

RESULT 11
AC110569/c
LOCUS AC110569 213457 bp DNA linear HTG 21-OCT-2002
DEFINITION Mus musculus clone RP23-309J17, WORKING DRAFT SEQUENCE, 7 ordered
pieces.
ACCESSION AC110569
VERSION AC110569.3 GI:24182091
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-309J17
Unpublished
2 (bases 1 to 213457)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McSwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 213457)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome

```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 21, 2002 this sequence version replaced gi:20503222.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L22620
Center clone name: 309 J.17

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 212308 bases at least Q40
Consensus quality: 212725 bases at least Q30
Consensus quality: 212828 bases at least Q20

Insert size: 210000; agarose-fp
Insert size: 212857; sum-of-contigs
Quality coverage: 11.3 in Q20 bases; agarose-fp
Quality coverage: 11.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1
65944: contig of 65944 bp in length
65945
66044: gap of 100 bp
66045
69682: contig of 3638 bp in length
69683
69783: gap of 100 bp
69784
73557: contig of 3875 bp in length
73558
73757: gap of 100 bp
73758
134380: contig of 60623 bp in length
134381
134481: gap of 100 bp
162491: contig of 28011 bp in length
162492
162591: gap of 100 bp
200818: contig of 38227 bp in length
200819
200919: gap of 100 bp
213457: contig of 12539 bp in length.

FEATURES

source
1. .213457
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RP23-309017"
/clone="RP23-309017"
/clone_lib="RP23-309017"
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
66045..69682
/notes="assembly_fragment"
69783..73557
/notes="assembly_fragment"
73758..134380
/notes="assembly_fragment"
134481..162491
/notes="assembly_fragment"
162592..200818
/notes="assembly_fragment"
200919..213457
/notes="assembly_fragment"
clone_end:T7
vector_side:right

ORIGIN

Query Match

Best Local Similarity 51.0%; Score 25; DB 2; Length 213457;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATAACACACCTTAGTCAGATACACTTT 41

DB 58385 GAATAAACACAGATTAGACCTTGTCATATTTACTTT 58345

RESULT 12

AL672023/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL672023 160615 bp DNA linear ROD 26-JUN-2002
Mouse DNA sequence from clone RP23-339019 on chromosome X, complete
sequence.
AL672023
AL672023.7 GI:21621646
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Tracey, A.
Direct Submission
Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 27, 2002 this sequence version replaced gi:21531223.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-339019 is
from the RPI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: PBACE3.6. Location/Qualifiers

FEATURES

source
1. .160615
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-339019"
/clone_lib="RP23-339019"

ORIGIN

Query Match 60.0%; Score 24.6; DB 10; Length 160615;
Best Local Similarity 76.9%; Pred. No. 35;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATAACACACCTTAGTCAGATACACTTT 39

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 18:56:32 ; Search time 214.674 Seconds
(without alignments)
811.351 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42

Perfect score: 41

Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactacttt 41

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003s.*

7: Geneseq2004s.*

8: Geneseq2005s.*

9: Geneseq2006s.*

10: Geneseq2007s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	1431	3	AAA71435 Human meg
2	40	97.6	128	3	AAA71434 Human meg
3	33.6	82.0	6182	6	ABL34015 Human inn
4	30	73.2	30	3	AAA71449 Human meg
5	30	73.2	30	3	AAA71448 Human meg
6	25	61.0	25	3	AAA71453 Human meg
7	25	61.0	6182	6	ABL34014 Human inn
8	24	58.5	5276	4	AAA46379 Human meg
9	24	58.5	5276	4	ABL32826 Human inn
10	23	56.1	23	3	AAA71452 Human meg
11	22.4	54.6	1380	9	ADC92907 E. faeciu
12	22.4	54.6	5204	6	ABL32899 Human inn
13	22.4	54.6	11052	6	ABL339985 Human che
14	22	53.7	6503	6	ABL32720 Human inn
15	21.8	53.2	2222	6	ABL34986 Human meg
16	21.8	53.2	3561	4	ABL13848 Drosophil
17	21.8	53.2	6568	6	AAA94755 Human DNA
18	21.6	52.7	1265	5	AA668177 Human lun
19	21.6	52.7	1265	6	ABL38088 cDNA enco
20	21.6	52.7	1265	7	ACAL0417 Human lun
21	21.6	52.7	1265	7	ABL39368 Lung canc
22	21.6	52.7	1265	10	ABE72151 Human lun
23	21.6	52.7	1350	5	AA872599 DNA enco

ALIGNMENTS

RESULT 1

AAA71435

ID AAA71435 standard; DNA; 1431 BP.

XX

AC AAA71435;

XX

DT 01-DEC-2000 (first entry)

XX

DE Human megsin promoter fragment DNA.

XX

KW Promoter; megsin; human; protein isolation; screening. ss.

XX

OS Homo sapiens.

XX

PN WO200043528-A1.

XX

PD 27-JUL-2000.

XX

PF 25-JAN-2000; 2000WO-JP000350.

XX

PR 25-JAN-1999; 99JP-00015667.

XX

PA (KURO/) KUROKAWA K.

XX

PI (MIYA/) MIYATA T.

XX

XX Miyata T;

XX

DR WPI; 2000-543257/49.

XX

PT DNA for promoter region of megsin useful for screening proteins.

XX

PS Disclosure; Fig 2; 45pp; Japanese.

XX

CC This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents a fragment of the human megsin promoter which is described in the method of the invention

XX

SQ Sequence 1431 BP; 466 A; 266 C; 303 G; 394 T; 0 U; 2 Other;

Query Match 100.0%; Score 41; DB 3; Length 1431;

Best Local Similarity 100.0%; Pred. No. 2.8e-06;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 41
 DB 1303 GAATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 1343

RESULT 2
 AAA71434
 ID AAA71434 standard; DNA; 128 BP.
 AC AAA71434;
 XX
 XX 01-DEC-2000 (first entry)
 DT
 DE Human meglin promoter fragment DNA.
 XX
 XX Promoter; meglin; human; protein isolation; screening. ss.
 KW Homo sapiens.
 OS
 XX WO200043528-A1.
 PN
 XX 27-JUL-2000.
 PD
 XX 25-JAN-2000; 2000WO-JP000350.
 PF
 XX 25-JAN-1999; 99JP-00015667.
 PR
 XX (KURO/) KUROKAWA K.
 PA (MIYA/) MIYATA T.
 XX
 XX Miyata T;
 PI
 XX WPI; 2000-543257/49.
 DR
 XX DNA for promoter region of meglin useful for screening proteins.
 PT
 XX Claim 1; Page 32; 45pp; Japanese.
 PS
 XX This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents the human meglin promoter which is described in the method of the invention
 CC
 XX Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
 SQ

Query Match 97.6%; Score 40; DB 3; Length 128;
 Best Local Similarity 100.0%; Pred. No. 4.6e-06;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 41
 DB 1 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 40

RESULT 3
 ABL34015/c
 ID ABL34015 standard; DNA; 6182 BP.
 AC ABL34015;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Human immune system associated gene SEQ ID NO: 1988.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosstatic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 DS.
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 XX 01-SEP-2000; 2000DE-01043826.
 XX
 XX (BPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 XX Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;
 SQ

Query Match 92.0%; Score 33.6; DB 6; Length 6182;
 Best Local Similarity 90.0%; Pred. No. 0.0029;
 Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 41
 DB 1310 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 1271

RESULT 4
 AAA71449
 ID AAA71449 standard; DNA; 30 BP.
 AC AAA71449;
 XX
 XX 01-DEC-2000 (first entry)
 DT
 XX Human meglin promoter PCR primer SEQ ID NO: 16.
 DE
 KW Promoter; meglin; human; protein isolation; screening. PCR primer; ss.
 KW Homo sapiens.
 OS
 XX WO200043528-A1.
 PN
 XX 27-JUL-2000.
 PD
 XX 25-JAN-2000; 2000WO-JP000350.
 PF
 XX 25-JAN-1999; 99JP-00015667.
 PR
 XX (KURO/) KUROKAWA K.
 PA (MIYA/) MIYATA T.
 XX
 XX Miyata T;
 PI

XX WPI; 2000-543257/49.
 XX DNA for promoter region of megins useful for screening proteins.
 XX Example 5; Page 40; 45pp; Japanese.
 XX This invention describes a novel DNA sequence (I) representing a promoter
 CC region having part or all of a specific base sequence. The invention also
 CC describes (1) a vector containing (I); (2) a cell transformed by the
 CC above vector; and (3) protein produced using (I). (I) is useful for
 CC screening and isolating proteins (especially transcription factors).
 CC AAA71434-A71469 represent PCR primers used in the method described in the
 CC invention
 XX Sequence 30 BP; 12 A; 9 C; 2 G; 7 T; 0 U; 0 Other;
 SQ
 Query Match 73.2%; Score 30; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.031; 0; Indels 0; Gaps 0;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 TACATAACCAACCACTTAGTCAGATAC 38
 Db 1 TACATAACCAACCACTTAGTCAGATAC 30
 RESULT 5
 AAA71448
 ID AAA71448 standard; DNA; 30 BP.
 XX AC AAA71448;
 XX DT 01-DEC-2000 (first entry)
 DE Human megsin promoter PCR primer SEQ ID NO: 15.
 DE Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
 KW Homo sapiens.
 XX WO200043528-A1.
 XX 27-JUL-2000.
 XX 25-JAN-2000; 2000WO-JP000350.
 XX 25-JAN-1999; 99JP-00015667.
 XX (KURO/) KUROKAWA K.
 XX (MIYA/) MIYATA T.
 XX Miyata T;
 WPI; 2000-543257/49.
 DNA for promoter region of megins useful for screening proteins.
 Example 5; Page 40; 45pp; Japanese.
 This invention describes a novel DNA sequence (I) representing a promoter
 region having part or all of a specific base sequence. The invention also
 describes (1) a vector containing (I); (2) a cell transformed by the
 above vector; and (3) protein produced using (I). (I) is useful for
 screening and isolating proteins (especially transcription factors).
 AAA71434-A71469 represent PCR primers used in the method described in the
 invention
 Sequence 30 BP; 13 A; 8 C; 3 G; 6 T; 0 U; 0 Other;
 Query Match 73.2%; Score 30; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.031;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAACCTACATAACCAACCACTTAGTCAG 31
 Db 1 AATGAACCTACATAACCAACCACTTAGTCAG 30
 RESULT 6
 AAA71453
 ID AAA71453 standard; DNA; 25 BP.
 XX AC AAA71453;
 XX DT 01-DEC-2000 (first entry)
 DE Human megsin promoter PCR primer SEQ ID NO: 20.
 DE Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
 KW Homo sapiens.
 XX WO200043528-A1.
 XX 27-JUL-2000.
 XX 25-JAN-2000; 2000WO-JP000350.
 XX 25-JAN-1999; 99JP-00015667.
 XX (KURO/) KUROKAWA K.
 XX (MIYA/) MIYATA T.
 XX Miyata T;
 WPI; 2000-543257/49.
 DNA for promoter region of megins useful for screening proteins.
 Example 4; Page 42; 45pp; Japanese.
 This invention describes a novel DNA sequence (I) representing a promoter
 region having part or all of a specific base sequence. The invention also
 describes (1) a vector containing (I); (2) a cell transformed by the
 above vector; and (3) protein produced using (I). (I) is useful for
 screening and isolating proteins (especially transcription factors).
 AAA71434-A71469 represent PCR primers used in the method described in the
 invention
 Sequence 25 BP; 8 A; 7 C; 2 G; 8 T; 0 U; 0 Other;
 Query Match 61.0%; Score 25; DB 3; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 17 AACCACTTAGTCAGATAC 41
 Db 1 AACCACTTAGTCAGATAC 25
 RESULT 7
 ABL34014
 ID ABL34014 standard; DNA; 6182 BP.
 XX AC ABL34014;
 XX DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 1987.
 DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1987; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 6182 BP; 1929 A; 37 C; 1260 G; 2956 T; 0 U; 0 Other;
XX
Query Match 61.0%; Score 25; DB 6; Length 6182;
Best Local Similarity 75.6%; Pred. No. 6.6;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 1 GAATGAATCATATACACACCTTAGTCAGATACACTTTT 41
DB 4872 GAATGAATATATATATATTTTATTAGTTAGTATATTATTT 4912
XX
RESULT 8
ID AAS46379/C
ID AAS46379 standard; DNA; 5276 BP.
XX
XX AAS46379;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #101.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002955.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX
XX 06-APR-2000; 2000DE-01019058.
XX
XX 07-APR-2000; 2000DE-01019173.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX

XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX
XX Claim 1; SEQ ID NO 101; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and oncogenes
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX 500 are missing from the sequence listing) sequences (Ss) and sequences
XX complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX probes for detecting the cytosine methylation state and/or single
XX nucleotide polymorphisms and also to be used in an array for analysing
XX diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX probes can also be used in a method for ascertaining genetic and/or
XX epigenetic parameters for the diagnosis and/or therapy of existing
XX diseases or the predisposition to specific diseases, by analysing
XX cytosine methylations. The parameters may be compared to another set of
XX genetic and/or epigenetic parameters, the differences serving as basis
XX for diagnosis and/or prognosis events which are disadvantageous to
XX patients. The present sequence is one of the 533 genomic sequences
XX derived from tumour suppressor genes and oncogenes. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 5276 BP; 1546 A; 88 C; 1102 G; 2540 T; 0 U; 0 Other;
XX
Query Match 58.5%; Score 24; DB 4; Length 5276;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 2 AATGACTACATACACACCTTAGTCAGATACACTTTT 41
DB 544 AATTACCTACATACACAACTCACTAACTCAAAATTCATTT 505
XX
RESULT 9
ABL32826/C
ID ABL32826 standard; DNA; 5276 BP.
XX
XX ABL32826;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 799.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX

```

PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 799; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 5276 BP; 1546 A; 88 C; 1102 G; 2540 T; 0 U; 0 Other;
XX
XX Query Match 58.5%; Score 24; DB 6; Length 5276;
XX Best Local Similarity 75.0%; Pred. No. 16;
XX Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0
XX
QY 2 AATGAACCTACATACACACCCCTTAGTCAGATACACTCTTT 41
   ||| ||||| ||||| ||| ||| ||||| ||
Db 544 AATTACCTACATACAACTCATAACTCAAAATTCACATT 505
   ||| ||||| ||||| ||| ||| ||||| ||

RESULT 10
AAAY1452
ID AAA71452 standard; DNA; 23 BP.
XX
XX AAA71452;
XX
XX 01-DEC-2000 (first entry)
XX
XX Human megsin promoter PCR primer SEQ ID NO: 19.
XX
XX Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
XX
XX Homo sapiens.
XX
XX WO200043528-A1.
XX
XX 27-JUL-2000.
XX
XX 25-JAN-2000; 2000WO-JP000350.
XX
XX 25-JAN-1999; 99JP-00015667.
XX
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX
XX Miyata T;
XX
XX WPI; 2000-543257/49.
XX
XX DNA for promoter region of megsin useful for screening proteins.
XX
XX Example 4; Page 42; 45pp; Japanese.
XX
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the
XX above vector; and (3) protein produced using (I). (I) is useful for
XX screening and isolating proteins (especially transcription factors).

```

```

CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
CC XX
SQ Sequence 23 BP; 11 A; 7 C; 2 G; 3 T; 0 U; 0 Other;

Query Match          56.1%; Score 23; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 GAATGAACACTACATAACAACACC 23
   |||||
DB 1 GAATGAACACTACATAACAACACC 23

RESULT 11
ADC92907/c
ID ADC92907 standard; DNA; 1380 BP.
XX
XX
AC ADC92907;
XX
XX DT
XX DT
XX DT
DE DE
XX DE
XX de; gene; urinary tract infection; bacteraemia; endocarditis; wound;
XX abdominal-pelvic infection.
XX
XX Enterococcus faecium.
OS
OS
PN US6583275-B1.
XX
XX 24-JUN-2003.
XX
XX 30-JUN-1998; 98US-00107532.
XX
XX 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2003-799836/75.
DR P-PSDB; ADC96561.
XX
XX
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX Example 1; SEQ ID NO 2534; 243pp; English.
XX
XX The invention relates to an isolated nucleic acid derived from
XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having
XX one of 10 fully defined sequences given in the (or comprising 40
XX sequential nucleotides chosen from any of the nucleic acids, its
XX complement or sequences hybridising to it). Also included are a
XX recombinant vector comprising the nucleic acid operably linked to
XX transcription regulatory element, a cell comprising the vector and a
XX single-stranded probe comprising the nucleic acid. The nucleic acids are
XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
XX The nucleic acids is useful for diagnosing pathological conditions
XX resulting from E. faecium bacterial infection (e.g. urinary tract
XX infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
XX infection) and for screening drugs such as agonists and antagonists. The
XX nucleic acid is useful for recombinant production of Candida albicans -
XX derived peptides or antisense polypeptides. Pharmaceutical compositions
XX and vaccines containing the nucleic acid are useful for preventing or
XX treating Enterococcus faecium infections. The present sequence represents
XX one if the disclosed E. faecium nucleic acids.
XX
XX Sequence 1380 BP; 475 A; 213 C; 306 G; 386 T; 0 U; 0 Other;
SQ

```

RESULT 13
ABK39985/c

RESULT 14
ABL32720/c

ID ABL32720 standard; DNA; 6503 BP.
 XX AC ABL32720;
 XX DT 26-MAR-2002 (first entry)
 XX DE Human immune system associated gene SEQ ID NO: 693.
 XX DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antieriosclerotic; antianaemic; cytosine; neotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX OS Homo sapiens.
 XX PN WO200200928-A2.
 XX PD 03-JAN-2002.
 XX PF 02-JUL-2001; 2001WO-BP007537.
 XX PR 30-JUN-2000; 2000DE-01032529.
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX DR WPI; 2002-130909/17.
 XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX PS Claim 1; SEQ ID NO 693; 32pp + Sequence Listing; German.
 XX CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX SQ Sequence 6503 BP; 1561 A; 85 C; 1446 G; 3411 T; 0 U; 0 Other;
 Query Match 53.7%; Score 22; DB 6; Length 6503;
 Best Local Similarity 73.7%; Pred. No. 1e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 2 AATGAACATACATACCAACCACTTAGTCAGATCTACT 39
 DB 509 AATAAACACATATAAATACCACTCGTCTTATTACT 472
 RESULT 15
 ABK34986/c
 ID ABK34986 standard; cDNA; 2222 BP.
 XX AC ABK34986;
 XX DT 08-MAY-2002 (first entry)
 XX DE Human cDNA encoding secreted protein #124.
 XX DE Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KW autoimmunity; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.
 XX OS Homo sapiens.
 XX PN WO200177288-A2.
 XX PD 18-OCT-2001.
 XX PF 29-MAR-2001; 2001WO-US010224.
 XX PR 06-APR-2000; 2000US-0195582P.
 XX PA (GEMV) GENETICS INST INC.
 XX PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 XX PI Gulukota K, Graham JR;
 XX DR WPI; 2002-179321/23.
 XX CC Five hundred and ninety two polynucleotides derived from a variety of
 CC human tissue sources which encode secreted proteins, useful for treating
 CC immune deficiencies and disorders such as autoimmune disorders.
 XX PS Claim 1; Page 130-131; 372pp; English.
 XX CC The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention
 XX SQ Sequence 2222 BP; 573 A; 581 C; 515 G; 553 T; 0 U; 0 Other;
 Query Match 53.2%; Score 21.8; DB 6; Length 2222;
 Best Local Similarity 70.7%; Pred. No. 1e+02;
 Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 1 GAATGACTACATACCAACCACTTAGTCAGATCTACTTT 41
 DB 2037 GACTGAACATACACCAAGGATCTTGTTCATAATCAATGT 1997
 Search completed: March 24, 2004, 22:28:26
 Job time : 217.674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:57:32 ; Search time 1996.1 Seconds
(without alignments)
613.370 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42

Perfect score: 41
Sequence: 1 gaatgaactacatacaacc.....ccttagtcagataactacttt 41

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gas_hum:*
18: em_gas_inv:*
19: em_gas_pin:*
20: em_gas_vrt:*
21: em_gas_fun:*
22: em_gas_mam:*
23: em_gas_mus:*
24: em_gas_pro:*
25: em_gas_rtd:*
26: em_gas_pug:*
27: em_gas_vrl:*
28: gb_gas1:*
29: gb_gas2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	58.5	523	13	BX499268
2	24	58.5	691	10	BS539741
3	24	58.5	713	14	CA788473
4	23.6	57.6	909	29	CNS0521Y

5	23.4	57.1	855	28	BX141244
6	23.2	56.6	991	29	CG156013
7	23.2	56.6	1018	29	CG431901
8	23.2	56.6	1054	29	CG431789
9	23	56.1	620	12	BJ624928
10	23	56.1	631	12	B096452
11	23	56.1	832	14	CA791463
12	22.8	55.6	517	28	AQ302725
13	22.8	55.6	542	13	BUT22189
14	22.8	55.6	725	28	A2896574
15	22.6	55.1	400	13	B224503
16	22.6	55.1	485	28	B43917
17	22.6	55.1	502	13	BQ104820
18	22.6	55.1	571	13	BX717912
19	22.6	55.1	584	9	AL656843
20	22.6	55.1	597	13	BQ395866
21	22.6	55.1	605	9	AL678053
22	22.6	55.1	623	9	AL677532
23	22.6	55.1	641	9	AL656534
24	22.6	55.1	643	13	BQ390368
25	22.6	55.1	651	9	AL634380
26	22.6	55.1	658	9	AL641573
27	22.6	55.1	660	9	AL662566
28	22.6	55.1	664	9	AL639562
29	22.6	55.1	676	10	BB087044
30	22.6	55.1	681	9	AL658689
31	22.6	55.1	776	14	CF148601
32	22.6	55.1	817	14	CF221876
33	22.6	55.1	861	13	BX715774
34	22.6	55.1	863	13	BX742495
35	22.6	55.1	891	14	CF240425
36	22.6	55.1	904	14	CF241852
37	22.6	55.1	935	14	CF220193
38	22.6	55.1	957	14	CF241149
39	22.6	55.1	1330	28	CC252939
40	22.6	55.1	1408	11	AK034567
41	22.6	55.1	260	12	BP020321
42	22.4	54.6	293	12	BP024827
43	22.4	54.6	304	10	BB269027
44	22.4	54.6	313	13	BW116259
45	22.4	54.6			

ALIGNMENTS

RESULT 1
BX499268
LOCUS
DEFINITION
523 bp mRNA linear EST 04-SEP-2003
DKFZp779I0544_r1 779 (synonym: hmccl) Homo sapiens cDNA clone
DKFZp779I0544_5', mRNA sequence.
ACCESSION
BX499268
VERSION
BX499268.1 GI:32017555
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 523)
AUTHORS
Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Oanger, A.,
Fob, G., Han, M. and Wiemann, S.
TITLE
EST (Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., et al.)
JOURNAL
Unpublished (2003)
COMMENT
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No si sequence available.

This clone (DKFZp779I0544) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
1. .523
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp779I0544"
/tissue_type="liver"
/dev_stage="fetal"
/lab_host="DH10B"
/clone_lib="779 (synonym: hmccl)"
/note="Vector: pSport1_Sfi; Site_1: SfiA; Site_2: SfiB"

ORIGIN

Query Match 58.5%; Score 24; DB 13; Length 523;
Best Local Similarity 75.0%; Pred. No. 3.1e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAATGAACACTACATACCAACCACTTGTAGTCAGATACTACTT 40
|||||
Db 137 GCATATACTAAATAGTTCACGCTAAGTCAGTACTACTT 176

RESULT 2

BE539741
LOCUS
DEFINITION
601061814F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3448443 5', mRNA sequence.
ACCESSION
BE539741
VERSION
BE539741.1 GI:9768386
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 691)
NIH-MGC http://mhc.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8423 row: k column: 04
High quality sequence start: 15
High quality sequence stop: 635.

FEATURES

source
1. .691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3448443"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC_10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5 kb. Library prepared by Life Technologies."

ORIGIN

Query Match 58.5%; Score 24; DB 10; Length 691;
Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAATGAACACTACATACCAACCACTTGTAGTCAGATACTACTT 40

Db

413 GCATATACTAAATAGTTCACGCTAAGTCAGTACTACTT 452

RESULT 3

CA788473
LOCUS
DEFINITION
AGENCOURT_10306216 NICHD XGC 001 Xenopus laevis cDNA clone IMAGE:5079161 5', mRNA sequence.
ACCESSION
CA788473
VERSION
CA788473.1 GI:26032557
KEYWORDS
EST.
SOURCE
Xenopus laevis (African clawed frog)
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 713)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11212 row: e column: 18
High quality sequence start: 89
High quality sequence stop: 288.

FEATURES

source
1. .713
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:5079161"
/tissue_type="oocytes"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC_001"
/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Life Technologies."

ORIGIN

Query Match 58.5%; Score 24; DB 14; Length 713;
Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACACTACATACCAACCACTTGTAGTCAGATACTACTT 41
|||||
Db 243 AATGAATAGATGACCAAGACCTTATATAGATCATATTT 282

RESULT 4

CNS0521Y
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone 003H09 of library B from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION
AL317599
VERSION
AL317599.1 GI:9550483
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 257 row: B column: 15
Seq primer: SP6
Class: BAC ends.

Location/Qualifiers
1. .855
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSNHed/MCW"
/db_xref="taxon:10116"
/clone="CH230-257B15"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/notes="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SSNHed/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 57.1%; Score 23.4; DB 28; Length 855;
Best Local Similarity 73.2%; Pred. No. 5.5e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 GAATGAACACTACATACACACCACTTGTAGTCAGACTACTCTTT 41
|||||
Db 636 GAATGAACACTAGAAACATCATCTCTGTAGTGAGGTAACACATT 676
|||||

RESULT 6

CG156013/c

LOCUS CG156013 991 bp DNA linear GSS 21-AUG-2003
DEFINITION PUIIP707D ZM 0.6-1.0 KB Zea mays genomic clone ZMMBTA0591L20,
genomic survey sequence.
ACCESSION CG156013
VERSION CG156013.1 GI:34046814
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 991)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reinick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUIIP70TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .991
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA0591L20"
/clone_lib="ZM 0.6-1.0 KB"
/notes="Vector: PCR4-TOP0; Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

```

Query Match          56.6%; Score 23.2; DB 29; Length 991;
Best Local Similarity 77.8%; Pred. No. 6.6e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AATGAATCATACACACCTTAGTCAGATACTA 37
Db 925 AGTATCTACAGAAAACACATAGTCAGATACTA 890

RESULT 7
CG431789/c
LOCUS CG431789 1018 bp DNA linear GSS 16-SEP-2003
DEFINITION Bg49-T7 Glomus intraradices Bgl II library (in pBK-CMV stratagene)
Glomus intraradices genomic clone Bg49-T7, genomic survey sequence.
ACCESSION CG431789
VERSION CG431789.1 GI:34766767
KEYWORDS GSS.
SOURCE Glomus intraradices
ORGANISM Glomus intraradices
Eukaryota; Fungi; Glomeromycota; Glomeromycetes; Glomerales;
Glomeraceae; Glomus.
REFERENCE 1 (bases 1 to 1018)
AUTHORS Lammers,P.J., Ratyanaka,S., Rehner,C. and Jun,J.
TITLE Genomic sequence from Glomus intraradices spore tissue
JOURNAL Unpublished (2003)
COMMENT Contact: Peter Lammers
New Mexico State University
BOX 30001, 3MLS, Las Cruces, NM 88003, USA
Tel: 505-646-6918
Fax: 505-646-6846
Email: plammers@nmsu.edu
Result of blastx search of NCBI non-redundant protein database,
June 2003: No significant matches.
Class: plasmid ends
Seq primer: T7
High quality sequence stop: 1018.
FEATURES
    source
        1..1018
            /organism="Glomus intraradices"
            /mol_type="genomic DNA"
            /strain="DAOM 197198, Biosystematic Research Center,
            Ottawa"
            /db_xref="taxon:4876"
            /clone="Bg49-T7"
            /tissue_type="Spores"
            /lab_host="E. coli XLR Blue"
            /clone_lib="Glomus intraradices Bgl II library (in pBK-CMV
            stratagene)"
            /note="Vector: pBK-CMV; genomic DNA isolated from Spores;
            complete Bgl II digest of gDNA; cloned into BamHI site of
            pBK-CMV"

ORIGIN
Query Match          56.6%; Score 23.2; DB 29; Length 1018;
Best Local Similarity 77.8%; Pred. No. 6.6e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AATGAATCATACACACCTTAGTCAGATACTA 37
Db 692 AATGGACACATACGATCACTTTGTGCACATAATA 657

RESULT 8
CG431789/c
LOCUS CG431789 1054 bp DNA linear GSS 16-SEP-2003
DEFINITION Bg110-T3 Glomus intraradices Bgl II library (in pBK-CMV stratagene)
Glomus intraradices genomic clone Bg110-T3, genomic survey
sequence.
ACCESSION CG431789
VERSION CG431789.1 GI:34766655
KEYWORDS GSS.
SOURCE Glomus intraradices
ORGANISM Glomus intraradices
Eukaryota; Fungi; Glomeromycota; Glomeromycetes; Glomerales;
Glomeraceae; Glomus.
REFERENCE 1 (bases 1 to 1054)
AUTHORS Lammers,P.J., Ratyanaka,S., Rehner,C. and Jun,J.
TITLE Genomic sequence from Glomus intraradices spore tissue
JOURNAL Unpublished (2003)
COMMENT Contact: Peter Lammers
New Mexico State University
BOX 30001, 3MLS, Las Cruces, NM 88003, USA
Tel: 505-646-6918
Fax: 505-646-6846
Email: plammers@nmsu.edu
Result of blastx search of NCBI non-redundant protein database,
June 2003: No significant matches.
Class: plasmid ends
Seq primer: T3
High quality sequence stop: 1054.
FEATURES
    source
        1..1054
            /organism="Glomus intraradices"
            /mol_type="genomic DNA"
            /strain="DAOM 197198, Biosystematic Research Center,
            Ottawa"
            /db_xref="taxon:4876"
            /clone="Bg110-T3"
            /tissue_type="Spores"
            /lab_host="E. coli XLR Blue"
            /clone_lib="Glomus intraradices Bgl II library (in pBK-CMV
            stratagene)"
            /note="Vector: pBK-CMV; genomic DNA isolated from Spores;
            complete Bgl II digest of gDNA; cloned into BamHI site of
            pBK-CMV"

ORIGIN
Query Match          56.6%; Score 23.2; DB 29; Length 1054;
Best Local Similarity 77.8%; Pred. No. 6.6e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AATGAATCATACACACCTTAGTCAGATACTA 37
Db 692 AATGGACACATACGATCACTTTGTGCACATAATA 657

RESULT 9
BJ624928
LOCUS BJ624928 620 bp mRNA linear EST 01-OCT-2003
DEFINITION Xenopus laevis cDNA clone XL210a01 5', mRNA sequence.
Xenopus laevis
ACCESSION BJ624928
VERSION BJ624928.1 GI:37269154
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 620)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
URL: http://xenopus.nibb.ac.jp.
The information of this clone is available through the following
URL.
FEATURES
    source
        1..620

```

```

Eukaryota; Fungi; Glomeromycota; Glomeromycetes; Glomerales;
Glomeraceae; Glomus.
REFERENCE 1 (bases 1 to 1054)
AUTHORS Lammers,P.J., Ratyanaka,S., Rehner,C. and Jun,J.
TITLE Genomic sequence from Glomus intraradices spore tissue
JOURNAL Unpublished (2003)
COMMENT Contact: Peter Lammers
New Mexico State University
BOX 30001, 3MLS, Las Cruces, NM 88003, USA
Tel: 505-646-6918
Fax: 505-646-6846
Email: plammers@nmsu.edu
Result of blastx search of NCBI non-redundant protein database,
June 2003: No significant matches.
Class: plasmid ends
Seq primer: T3
High quality sequence stop: 1054.
FEATURES
    source
        1..1054
            /organism="Glomus intraradices"
            /mol_type="genomic DNA"
            /strain="DAOM 197198, Biosystematic Research Center,
            Ottawa"
            /db_xref="taxon:4876"
            /clone="Bg110-T3"
            /tissue_type="Spores"
            /lab_host="E. coli XLR Blue"
            /clone_lib="Glomus intraradices Bgl II library (in pBK-CMV
            stratagene)"
            /note="Vector: pBK-CMV; genomic DNA isolated from Spores;
            complete Bgl II digest of gDNA; cloned into BamHI site of
            pBK-CMV"

ORIGIN
Query Match          56.6%; Score 23.2; DB 29; Length 1054;
Best Local Similarity 77.8%; Pred. No. 6.6e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AATGAATCATACACACCTTAGTCAGATACTA 37
Db 692 AATGGACACATACGATCACTTTGTGCACATAATA 657

RESULT 9
BJ624928
LOCUS BJ624928 620 bp mRNA linear EST 01-OCT-2003
DEFINITION Xenopus laevis cDNA clone XL210a01 5', mRNA sequence.
Xenopus laevis
ACCESSION BJ624928
VERSION BJ624928.1 GI:37269154
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 620)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
URL: http://xenopus.nibb.ac.jp.
The information of this clone is available through the following
URL.
FEATURES
    source
        1..620

```

```

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL210a01"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula
library"

ORIGIN
Query Match          56.1%; Score 23; DB 12; Length 620;
Best Local Similarity 74.4%; Pred. No. 7.2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACATCATACCAACCACTTAGTCAGATCACTACTT 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 AATGAATAGATGACAAAGACCTTATATAGATCATTAATT 379

RESULT 10
BJ096452          631 bp mRNA linear EST 01-OCT-2003
LOCUS
DEFINITION
Xenopus laevis cdna clone XL154a01 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 631)
AUTHORS
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
Expressed genes in X. laevis embryo
Contact: Tadasu Shin-i
Unpublished (2001)
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshinigenes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.

FEATURES
source
Location/Qualifiers
1..631
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL154a01"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula
library"

ORIGIN
Query Match          56.1%; Score 23; DB 12; Length 631;
Best Local Similarity 74.4%; Pred. No. 7.2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACATCATACCAACCACTTAGTCAGATCACTACTT 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 AATGAATAGATGACAAAGACCTTATATAGATCATTAATT 588

RESULT 11
CA791463          832 bp mRNA linear EST 04-DEC-2002
LOCUS
DEFINITION
AGENCOURT 10304580 NICHG_XGC_Emb1 Xenopus laevis cdna clone
IMAGE:5161647 5', mRNA sequence.

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 832)
REFERENCE
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11401 row: b column: 16
High quality sequence stop: 417.
Location/Qualifiers
1..832
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:5161647"
/tissue_type="embryo (stage 10)"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHG XGC Emb1"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 1.55 kb. Constructed by Life Technologies. Note: This
is a Xenopus Gene Collection (XGC) library."

ORIGIN
Query Match          56.1%; Score 23; DB 14; Length 832;
Best Local Similarity 74.4%; Pred. No. 7.5e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACATCATACCAACCACTTAGTCAGATCACTACTT 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 AATGAATAGATGACAAAGACCTTATATAGATCATTAATT 191

RESULT 12
AQ302725          517 bp DNA linear GSS 16-DEC-1998
LOCUS
DEFINITION
HS 3217 B1_C10 MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3217 Col=19 Row=F, genomic survey
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 517)
REFERENCE
AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
PUBMED
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington

```

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3217 row: F column: 19

Class: BAC ends

High quality sequence stop: 517.

Location/Qualifiers

1. 517

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clones="Plate:3217 Col=19 Row=F"

/sex="male"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

ORIGIN

Query Match 55.6%; Score 22.8; DB 28; Length 517;
Best Local Similarity 79.4%; Pred. No. 8.2e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 7 ACTACATAACACACCTTAGTCAGATACTACTT 40

Db 280 ACGACGTGAAGAACCATGTGAGGCAGATACTACTT 313

RESULT 13

LOCUS

DEFINITION BU722189 542 bp mRNA linear EST 23-OCT-2003

ACCESSION SJNABF03 Schistosoma Japonicum cDNA, mRNA sequence.

VERSION BU722189

KEYWORDS BU722189.1 GI:28329558

SOURCE EST

ORGANISM Schistosoma japonicum

Schistosoma japonicum

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigoida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1. (bases 1 to 542)

Hu, W., Yan, Q., Shen, D. X., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R.,

Wang, Z. J., Kong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J.,

Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J.,

McManus, D. P., Xue, C. L., Feng, Z., Chen, Z. and Han, Z. G.

Evolutionary and biomedical implications of a Schistosoma japonicum

complementary DNA resource

Nat. Genet. 35 (2), 139-147 (2003)

22879925

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn.

Location/Qualifiers

1. 542

/organism="Schistosoma japonicum"

/mol_type="mRNA"

/db_xref="taxon:6182"

/sex="male"

/tissue_type="Whole body"

/dev_stage="adult"

/lab_host="rabbits"

/clone_lib="SJM"

ORIGIN

Query Match 55.6%; Score 22.8; DB 13; Length 542;
Best Local Similarity 79.4%; Pred. No. 8.3e+02;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 6 AACTACATAACACACCTTAGTCAGATACTACTT 39

Db 119 ATCTACATAAGAACCCCTAAATAAGATAACT 152

RESULT 14

LOCUS

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 00:16:57 ; Search time 1704 Seconds
(without alignments)
1042.878 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42
Perfect score: 41
Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactacttt 41

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sta.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sta.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	37	90.2	4229	9	AF234618	AF234618 Homo sapi	
c	2	37	90.2	157284	9	AC072051	AC072051 Homo sapi
3	37	90.2	189092	2	AC027524	AC027524 Homo sapi	
c	4	21	51.2	6182	6	AX346917	AX346917 Sequence
c	5	19	46.3	37490	9	AC026356	AC026356 Homo sapi
6	19	46.3	74350	8	AP002543	AP002543 Arabidops	
7	19	46.3	156394	2	AC026558	AC026558 Homo sapi	
8	19	46.3	220480	2	AC023973	AC023973 Homo sapi	
9	18	43.9	2000	6	AX655811	AX655811 Sequence	
c	10	18	43.9	143961	2	AP005862	AP005862 Oryza sat
11	18	43.9	159056	8	CNS08CD7	AL954853 Oryza sat	
12	18	43.9	215960	2	AC125959	AC125959 Rattus no	
13	18	43.9	282132	2	AC109976	AC109976 Rattus no	
c	14	17	41.5	1712	8	AK073122	AK073122 Oryza sat
c	15	17	41.5	16918	6	AX346519	AX346519 Sequence
c	16	17	41.5	107365	4	AC087861	AC087861 Felis cat
c	17	17	41.5	114149	9	AC022083	AC022083 Homo sapi
c	18	17	41.5	117995	10	AL808138	AL808138 Mouse DNA
c	19	17	41.5	135455	9	AL512666	AL512666 Human DNA
20	17	17	41.5	148510	10	AC113889	AC113889 Rattus no
21	17	17	41.5	152560	9	AC011339	AC011339 Homo sapi
22	17	17	41.5	157585	2	BX537352	BX537352 Mus muscu
c	23	17	41.5	162215	10	AL731773	AL731773 Mouse DNA
c	24	17	41.5	162621	2	AC145531	AC145531 Lemur cat
25	17	41.5	165764	9	AC116917	AC116917 Homo sapi	
26	17	41.5	173933	9	AP002022	AP002022 Homo sapi	
27	17	41.5	180905	2	AC011688	AC011688 Homo sapi	
28	17	41.5	188818	2	AC098674	AC098674 Homo sapi	
29	17	41.5	189715	2	AC025558	AC025558 Homo sapi	
30	17	41.5	205202	2	AC139347	AC139347 Mus muscu	
31	17	41.5	222932	2	AC137869	AC137869 Mus muscu	
c	32	17	41.5	231197	2	AC128935	AC128935 Rattus no
c	33	17	41.5	234545	5	BX470214	BX470214 Zebrafish
c	34	17	41.5	269132	2	AC118381	AC118381 Rattus no
35	17	41.5	293576	2	AC097830	AC097830 Rattus no	
36	16	39.0	1766	3	AF070917	AF070917 Drosophil	
c	37	16	39.0	2076	3	AF429950	AF429950 Plasmodiu
c	38	16	39.0	4116	8	SS1460	SS1460 patatin A g
c	39	16	39.0	6160	6	AX344268	AX344268 Sequence
c	40	16	39.0	6160	6	AX348665	AX348665 Sequence
c	41	16	39.0	17993	6	AX711896	AX711896 Sequence
c	42	16	39.0	19914	2	AC018009	AC018009 Drosophil
c	43	16	39.0	37650	9	HSU131B10	Z73417 Human DNA s
c	44	16	39.0	51680	9	AP000261	AP000261 Homo sapi
45	16	39.0	54488	8	AF359360	AF359360 Fusarium	

ALIGNMENTS

RESULT 1
AF234618
LOCUS AF234618 Homo sapiens megin gene, promoter region and partial sequence.
DEFINITION AF234618 Homo sapiens megin gene, promoter region and partial sequence.
ACCESSION AF234618.2 GI:19808130
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4229)
AUTHORS Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
TITLE Transcriptional regulation of a mesangium-predominant gene, megin
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4229)
 AUTHORS Miyata,T., Nangaku,M., Inagi,R. and Kurokawa,K.
 TITLE Direct Submission
 JOURNAL Submitted (15-FEB-2000) Institute of Medical Sciences and
 Department of Internal Medicine, Tokai University School of
 Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan

REFERENCE 3 (bases 1 to 4229)
 AUTHORS Miyata,T., Nangaku,M., Inagi,R. and Kurokawa,K.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2002) Institute of Medical Sciences and
 Department of Internal Medicine, Tokai University School of
 Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan

REMARK
 COMMENT Sequence update by submitter
 On Mar 29, 2002 this sequence version replaced gi:18000453.

FEATURES
 source
 1..4229
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q21.3"
 1..4021
 4022..>4229
 /product="megsin"

promoter
 mRNA

ORIGIN
 Query Match 90.2% Score 37; DB 9; Length 4229;
 Best Local Similarity 100.0%; Pred. No. 4.6e-12;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATGAACATACATACACACCTTAGTCAGATACTA 37
 |||||||
 Db 3920 GAATGAACATACATACACACCTTAGTCAGATACTA 3956

RESULT 2
 AC072051/c
 LOCUS 157284 bp DNA linear PRI 03-DEC-2001
 DEFINITION Homo sapiens chromosome , clone RP11-79D21, complete sequence.
 AC072051
 ACCESSION AC072051.8 GI:16974280
 VERSION HTG.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 157284)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,
 Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
 Gage,D., Galagan,J., Gage,D., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gordon,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hags,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McDonald,P., Marquis,N.,
 Meldrim,J., Meneus,L., Mihova,T., Mhova,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,R., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

REFERENCE 2 (bases 1 to 157284)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,
 Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
 Gage,D., Galagan,J., Gage,D., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gordon,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hags,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McDonald,P., Marquis,N.,
 Meldrim,J., Meneus,L., Mihova,T., Mhova,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,R., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

REFERENCE 3 (bases 1 to 157284)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,
 Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
 Gage,D., Galagan,J., Gage,D., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gordon,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hags,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McDonald,P., Marquis,N.,
 Meldrim,J., Meneus,L., Mihova,T., Mhova,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,R., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

JOURNAL
 REFERENCE
 3 (bases 1 to 157284)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hags,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mhova,T., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,R., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL
 REFERENCE
 4 (bases 1 to 157284)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hags,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., McKernan,K., McPheeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Mhova,T., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,R., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (03-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT
 On Nov 18, 2001 this sequence version replaced gi:14277307.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L966
 Center clone name: 79_D_21

----- Location/Qualifiers
 1..157284
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-79D21"

FEATURES
 source
 1..157284

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzGerald, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Labrecque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneses, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2002 this sequence version replaced gi:11990731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8516
Center clone name: 317_0_1

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 85659: contig of 85659 bp in length
* 85660 85759: gap of 100 bp
* 85760 163014: contig of 77255 bp in length
* 163015 163114: gap of 100 bp
* 163115 189092: contig of 25978 bp in length.

FEATURES
source
1. 189092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="18"
/map="18"
/clone="RP11-317G1"
/clone_lib="RP11-317G1 Human Male BAC"

ORIGIN

Query Match 90.2%; Score 37; DB 2; Length 189092;
Best Local Similarity 100.0%; Pred. No. 2.9e-12;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAACATACATAACACCACTTACTGATGATCTA 37
|||||

Db 66563 GAATGAACATACATAACACCACTTACTGATGATCTA 66599
|||||

RESULT 4

AX346917/c
LOCUS
DEFINITION Sequence 1988 from Patent WO0200928.
ACCESSION AX346917

VERSION AX346917.1 GI:18494803
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1988 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
1. 6182
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)"
ORIGIN
Query Match 51.2%; Score 21; DB 6; Length 6182;
Best Local Similarity 100.0%; Pred. No. 0.038; 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;
QY 6 AACTACATAACACCACTTCTA 26
|||||
Db 1306 AACTACATAACACCACTTCTA 1286
|||||
RESULT 5
LOCUS AC026356/c
DEFINITION Homo sapiens 12 BAC RP11-81714 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC026356
VERSION AC026356.29 GI:15789206
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Muzny, D.M., Adams, C., Altschuler, H.C., Are, J.R., Banks, T., Barbieri, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowe, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marandel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Oguh, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pull, L., Quiles, M., Ren, Y., Rives, M., Rojao, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shim, C.,


```

/evidence=not experimental
/product="ABC transporter-like protein"
/protein_id="BAB11402.1"
/db_xref="GI:10178109"
/translation="MNQCNVTVSABEDIBAGKKPKQAEPTLPFLFKRDVTVKVI
KGLTSVEKILITSGSVNPGVSLMPSGSGKTLISLAGRISQSTGSGSVYN
DKPYKSLKSGIFVQDDVLPFLHVTARLRLPKTLTRBQKQALDVTQIE
LGLERCQDTMGAGFVGVGGGRKRVISGNEIINPSLLLDDEPTSGLDSTALATI
LMLDLAEAGKTVITTHQPSRLFRFPDKLILGRSLLYFGKSEALDYFSSIGCS
PLIAMPAPQLLDLANGINDISVSELDORVQNSGRQTKGKSPAAVHEYLVEA
YETRAEQKKLLDLPVDEEAKAKSTRLKRWGQTCWEEQICLFCGLKERRHEVF
SWLRVTVLSLAVILGLWQSDIRTPMGLQDQAGLLFTIAVFWGFFVFTAFAPPO
ERAMLNKRAADMYRLSAYFLARTSDLPDLPFLPSLFLVIVFMTGLAISVPYFPLS
MLTVFLCIIIAOQGLAIGAILMDLKKATTLASVTVMFLMAGGFVFKASPLFLDLC
P"
complement (32784..25652)
/notes="gene_id:F15M7.7"
/codon_start=1
/evidence=not experimental
/product="selenium-binding protein-like"
/protein_id="BAB11403.1"
/db_xref="GI:10178110"
/translation="MENVILNTRFKPKLALLQSCSPSDDLKIHGFLRLTHLSDV
FVSRLLALCVDDSTNFKPTNLGLYAGIFSOIQNPFLNLLRCFCSTGAPSRFAF
GYTOMLKSRIVPDNTTFPFLIKASMECVLAGEQTHSQIVRFGQNDVYVENSLSVH
MYANGCFIAAGRIPOGFRDVSMTVMVAGYCKGVENAREMDFEMPHRLFTWS
IMINGYAKNCEKADLPEFKRGKGVANETVMUSVISCAHLGALGEGRAYEVV
KEHMTVNLILGALVDMFRCQDIEKAIHVFEGLPDTSLSSWSSILKGLAVGHAKHA
MYHFSQISGLFPDVTVALSACSHGLVEKGLIYENMKKHGIEPRLEHYGCI
VMDLGRAGLAEENILKRWKPNAPILGALGACKIYKTEVAERVGNNMLIKVKE
HGYVYLLSNVYACQWQDKIESLRDMKELVPPKPGWSLJEDIKNFKTPDAGDQK
HPGMYKRKWEILGKRLIGYKNGTDAFPDVBDEEKESLTHMSEKLAITYAGMMK
TKPGTIRIVKLRVCECHVTKLISVYVRELIVRDRNRFFHFGVSCRDY"
complement (26394..27671)
/notes="contains similarity to unknown protein
gb|AA071807.1
gene_id:F15M7.8"
/codon_start=1
/evidence=not experimental
/protein_id="BAB11404.1"
/db_xref="GI:10178111"
/translation="MPKCNLLTTSKRRKSKRLKHQHPESLPPEKEVEEDRDE
GGFKLIAAPSGHEQVQPNLYFNPGAVNRTGLGNQILLSDELVIDLILGLGANH
LGVLAIVTKSFYIPANHEPLMNLVLEELKGLDFLNGSKRSTVVAHYHKKFKPAGDGE
SNLKIIDVSYDLFQSMCANLEMKPKLRRONITVRGIVSEDFITKEEPKPKVLL
ESCLDQWPALEKWSRDLKLVNVDVEFAVGPVEMKLEKFPKVSQDAREPRLYLEDPK
FAEKVPLSEVDVYVFRDLFGVGNRKPDRWIIIGPAGSGSFHIDPDNSTAWN
AVITGSKWLPFPDVPVPGVHPSPDGAECVPSIENFMNPFYDDTKDWEKKPIECI
CKAGEVFPVNGWHLVINLESIAITQYASR"
complement (Join(28281..28383,28556..28602,28740..30146))
/notes="gb|AA023201.1
gene_id:F15M7.9
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB11405.1"
/db_xref="GI:10178112"
/translation="MDLATVSPTRDVRCCDGLNGASPGSLLSVSKRKYEEF
ENEKLFHIELELDLSNAKVQLENELELLRETSSQQQSIQDLFELDEERNAASTA
ASRAMNLRQRKQAKQELQKFAFERKMEHQDQELDLLELIIYKREQTIQALT
FQAQYKHMRSFGFEAEVETEKNNLSPNMIENDYQDLDTSYDPIKQVNNP
GLEADIDVDVVEKYLADSPHLKTLERRISQMERNPSTQPTQDVGSGRHYTKV
VQSPRQHQHFRVRSYSSSLGTTREKRLDPSNPSRNSNGSPKSGRHYTKV
FARDKGDSEIIGNMDRVYITDSVHSHVSHSGTAEQKPKNDADGYAMSRELINQ
PLGDPEISKLRYNLQALREDSMQAITSRRTEKXQMWLLKEIAQLHLSKDVVPERR
LPURKTSIIIGAFNFSVFKWISFVFRKRRSKYMGVQGNMGLQMLLEKTPRIR
QWRCSSTQV"
complement (Join(32678..32984,34934..35616))
/notes="gb|AAD04946.2
gene_id:F15M7.10
similar to unknown protein"
/codon_start=1
/evidence=not experimental

```

CDS

CDS

CDS

CDS

```

/protein_id="BAB11406.1"
/db_xref="GI:10178113"
/translation="MGSLEEQVAEDCMGLIQLLSNGVTLRSESDILITQIIPFRNN
Query Match 46.3%; Score 19; DB 8; Length 74350;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 CAACCACTTAGTCAGATA 34
|||||
Db 26951 CAACCACTTAGTCAGATA 26969
|||||
RESULT 7
AC026558 156394 bp DNA linear HTG 26-MAY-2000
LOCUS Homo sapiens clone RP11-637N6, WORKING DRAFT SEQUENCE, 23 unordered
pieces.
AC026558 3 GI:8077003
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156394)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE Homo sapiens, clone RP11-637N6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156394)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhaltier,B., Brown,A., Burkett,G.,
Campolano,A., Castele,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,N., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7523786.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7470
Center clone name: 637 N 6
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 144229 bases at least Q40
Consensus quality: 150107 bases at least Q30
Consensus quality: 152252 bases at least Q20

```

Insert size: 163000; agarose-fp
 Insert size: 154194; sum-of-contigs
 Quality coverage: 4.2 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1014: contig of 1014 bp in length
 * 1015 1114: gap of 100 bp
 * 1115 2440: contig of 1326 bp in length
 * 2441 2540: gap of 100 bp
 * 2541 4502: contig of 1962 bp in length
 * 4503 4602: gap of 100 bp
 * 4603 5920: contig of 1318 bp in length
 * 5921 6020: gap of 100 bp
 * 6021 8010: contig of 1990 bp in length
 * 8011 9645: contig of 1535 bp in length
 * 9646 9745: gap of 100 bp
 * 9746 11265: contig of 1520 bp in length
 * 11266 12378: gap of 100 bp
 * 12379 12478: gap of 100 bp
 * 12479 14550: contig of 2072 bp in length
 * 14551 14650: gap of 100 bp
 * 14651 17260: contig of 2610 bp in length
 * 17261 17360: gap of 100 bp
 * 17361 19770: contig of 2410 bp in length
 * 19771 19870: gap of 100 bp
 * 19871 24405: contig of 4535 bp in length
 * 24406 24505: gap of 100 bp
 * 24506 28007: contig of 3502 bp in length
 * 28008 28107: gap of 100 bp
 * 28108 32250: contig of 4143 bp in length
 * 32251 32350: gap of 100 bp
 * 32351 38617: contig of 6267 bp in length
 * 38618 38717: gap of 100 bp
 * 38718 46909: contig of 8192 bp in length
 * 46910 47010: gap of 100 bp
 * 47011 56251: contig of 9242 bp in length
 * 56252 56351: gap of 100 bp
 * 56352 64665: contig of 8314 bp in length
 * 64666 64765: gap of 100 bp
 * 64766 77790: contig of 13025 bp in length
 * 77791 77990: gap of 100 bp
 * 77991 88643: contig of 10753 bp in length
 * 88644 88743: gap of 100 bp
 * 88744 102574: contig of 13831 bp in length
 * 102575 102674: gap of 100 bp
 * 102675 122629: contig of 19955 bp in length
 * 122630 122729: gap of 100 bp
 * 122730 156394: contig of 33665 bp in length.

FEATURES

source
 1. .156394
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clones="RP11-637M6"
 /clone_lib="RP11-11 Human Male BAC"
 1. .1014
 /note="assembly_fragment"
 1115..2440
 /note="assembly_fragment"
 2541..4502
 /note="assembly_fragment"
 4603..5920
 /note="assembly_fragment"

misc_feature 6021..8010
 /note="assembly_fragment"
 misc_feature 8111..9645
 /note="assembly_fragment"
 misc_feature 9746..11265
 /note="assembly_fragment"
 misc_feature 11366..12378
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 12479..14550
 /note="assembly_fragment"
 misc_feature 14651..17260
 /note="assembly_fragment"
 misc_feature 17361..19770
 /note="assembly_fragment"
 misc_feature 19871..24405
 /note="assembly_fragment"
 misc_feature 24506..28007
 /note="assembly_fragment"
 misc_feature 28108..32250
 /note="assembly_fragment"
 misc_feature 32351..38617
 /note="assembly_fragment"
 misc_feature 38718..46909
 /note="assembly_fragment"
 misc_feature 47010..56251
 /note="assembly_fragment"
 misc_feature 56352..64665
 /note="assembly_fragment"
 misc_feature 64766..77790
 /note="assembly_fragment"
 misc_feature 77891..88643
 /note="assembly_fragment"
 misc_feature 88744..102574
 /note="assembly_fragment"
 misc_feature 102675..122629
 /note="assembly_fragment"
 misc_feature 122730..156394
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left"

ORIGIN

Query Match 46.3%; Score 19; DB 2; Length 156394;
 Best Local Similarity 100.0%; Pred. No. 0.44;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AACTACATACACACCT 24
 |||||
 Db 48582 AACTACATACACACCT 48600

RESULT 8

AC023973 220480 bp DNA linear HTG 23-SEP-2000
 Homo sapiens chromosome 12 clone RP11-793A9 map 12, WORKING DRAFT
 SEQUENCE, 28 unordered pieces.
 AC023973
 AC023973.4 GI:10280873
 HTG; HTGS PHASE1; HTGS DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 220480)
 Birren,B., Linton,L., Nusbaum,C. and Lander,B.
 Homo sapiens chromosome 12, clone RP11-793A9
 Unpublished
 2 (bases 1 to 220480)
 Birren,B., Linton,L., Nusbaum,C., Lander,B., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Chospel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Feneator, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Sierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Largocque, K., Leloczky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Minova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivat, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8076863.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----
 Contact: sequence_submission@genome.wi.mit.edu
 Center project name: 16590

Center project name: L0580
Center clone name: 793 A 9

```
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemical structure: C1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3C4=CC=CC=C4C5=CC=CC=C5C6=CC=CC=C6C7=CC=CC=C7C8=CC=CC=C8C9=CC=CC=C9C10=CC=CC=C10C11=CC=CC=C11C12=CC=CC=C12C13=CC=CC=C13C14=CC=CC=C14C15=CC=CC=C15C16=CC=CC=C16C17=CC=CC=C17C18=CC=CC=C18C19=CC=CC=C19C20=CC=CC=C20C21=CC=CC=C21C22=CC=CC=C22C23=CC=CC=C23C24=CC=CC=C24C25=CC=CC=C25C26=CC=CC=C26C27=CC=CC=C27C28=CC=CC=C28C29=CC=CC=C29C30=CC=CC=C30C31=CC=CC=C31C32=CC=CC=C32C33=CC=CC=C33C34=CC=CC=C34C35=CC=CC=C35C36=CC=CC=C36C37=CC=CC=C37C38=CC=CC=C38C39=CC=CC=C39C40=CC=CC=C40C41=CC=CC=C41C42=CC=CC=C42C43=CC=CC=C43C44=CC=CC=C44C45=CC=CC=C45C46=CC=CC=C46C47=CC=CC=C47C48=CC=CC=C48C49=CC=CC=C49C50=CC=CC=C50C51=CC=CC=C51C52=CC=CC=C52C53=CC=CC=C53C54=CC=CC=C54C55=CC=CC=C55C56=CC=CC=C56C57=CC=CC=C57C58=CC=CC=C58C59=CC=CC=C59C60=CC=CC=C60C61=CC=CC=C61C62=CC=CC=C62C63=CC=CC=C63C64=CC=CC=C64C65=CC=CC=C65C66=CC=CC=C66C67=CC=CC=C67C68=CC=CC=C68C69=CC=CC=C69C70=CC=CC=C70C71=CC=CC=C71C72=CC=CC=C72C73=CC=CC=C73C74=CC=CC=C74C75=CC=CC=C75C76=CC=CC=C76C77=CC=CC=C77C78=CC=CC=C78C79=CC=CC=C79C80=CC=CC=C80C81=CC=CC=C81C82=CC=CC=C82C83=CC=CC=C83C84=CC=CC=C84C85=CC=CC=C85C86=CC=CC=C86C87=CC=CC=C87C88=CC=CC=C88C89=CC=CC=C89C90=CC=CC=C90C91=CC=CC=C91C92=CC=CC=C92C93=CC=CC=C93C94=CC=CC=C94C95=CC=CC=C95C96=CC=CC=C96C97=CC=CC=C97C98=CC=CC=C98C99=CC=CC=C99C100=CC=CC=C100C101=CC=CC=C101C102=CC=CC=C102C103=CC=CC=C103C104=CC=CC=C104C105=CC=CC=C105C106=CC=CC=C106C107=CC=CC=C107C108=CC=CC=C108C109=CC=CC=C109C110=CC=CC=C110C111=CC=CC=C111C112=CC=CC=C112C113=CC=CC=C113C114=CC=CC=C114C115=CC=CC=C115C116=CC=CC=C116C117=CC=CC=C117C118=CC=CC=C118C119=CC=CC=C119C120=CC=CC=C120C121=CC=CC=C121C122=CC=CC=C122C123=CC=CC=C123C124=CC=CC=C124C125=CC=CC=C125C126=CC=CC=C126C127=CC=CC=C127C128=CC=CC=C128C129=CC=CC=C129C130=CC=CC=C130C131=CC=CC=C131C132=CC=CC=C132C133=CC=CC=C133C134=CC=CC=C134C135=CC=CC=C135C136=CC=CC=C136C137=CC=CC=C137C138=CC=CC=C138C139=CC=CC=C139C140=CC=CC=C140C141=CC=CC=C141C142=CC=CC=C142C143=CC=CC=C143C144=CC=CC=C144C145=CC=CC=C145C146=CC=CC=C146C147=CC=CC=C147C148=CC=CC=C148C149=CC=CC=C149C150=CC=CC=C150C151=CC=CC=C151C152=CC=CC=C152C153=CC=CC=C153C154=CC=CC=C154C155=CC=CC=C155C156=CC=CC=C156C157=CC=CC=C157C158=CC=CC=C158C159=CC=CC=C159C160=CC=CC=C160C161=CC=CC=C161C162=CC=CC=C162C163=CC=CC=C163C164=CC=CC=C164C165=CC=CC=C165C166=CC=CC=C166C167=CC=CC=C167C168=CC=CC=C168C169=CC=CC=C169C170=CC=CC=C170C171=CC=CC=C171C172=CC=CC=C172C173=CC=CC=C173C174=CC=CC=C174C175=CC=CC=C175C176=CC=CC=C176C177=CC=CC=C177C178=CC=CC=C178C179=CC=CC=C179C180=CC=CC=C180C181=CC=CC=C181C182=CC=CC=C182C183=CC=CC=C183C184=CC=CC=C184C185=CC=CC=C185C186=CC=CC=C186C187=CC=CC=C187C188=CC=CC=C188C189=CC=CC=C189C190=CC=CC=C190C191=CC=CC=C191C192=CC=CC=C192C193=CC=CC=C193C194=CC=CC=C194C195=CC=CC=C195C196=CC=CC=C196C197=CC=CC=C197C198=CC=CC=C198C199=CC=CC=C199C200=CC=CC=C200C201=CC=CC=C201C202=CC=CC=C202C203=CC=CC=C203C204=CC=CC=C204C205=CC=CC=C205C206=CC=CC=C206C207=CC=CC=C207C208=CC=CC=C208C209=CC=CC=C209C210=CC=CC=C210C211=CC=CC=C211C212=CC=CC=C212C213=CC=CC=C213C214=CC=CC=C214C215=CC=CC=C215C216=CC=CC=C216C217=CC=CC=C217C218=CC=CC=C218C219=CC=CC=C219C220=CC=CC=C220C221=CC=CC=C221C222=CC=CC=C222C223=CC=CC=C223C224=CC=CC=C224C225=CC=CC=C225C226=CC=CC=C226C227=CC=CC=C227C228=CC=CC=C228C229=CC=CC=C229C230=CC=CC=C230C231=CC=CC=C231C232=CC=CC=C232C233=CC=CC=C233C234=CC=CC=C234C235=CC=CC=C235C236=CC=CC=C236C237=CC=CC=C237C238=CC=CC=C238C239=CC=CC=C239C240=CC=CC=C240C241=CC=CC=C241C242=CC=CC=C242C243=CC=CC=C243C244=CC=CC=C244C245=CC=CC=C245C246=CC=CC=C246C247=CC=CC=C247C248=CC=CC=C248C249=CC=CC=C249C250=CC=CC=C250C251=CC=CC=C251C252=CC=CC=C252C253=CC=CC=C253C254=CC=CC=C254C255=CC=CC=C255C256=CC=CC=C256C257=CC=CC=C257C258=CC=CC=C258C259=CC=CC=C259C260=CC=CC=C260C261=CC=CC=C261C262=CC=CC=C262C263=CC=CC=C263C264=CC=CC=C264C265=CC=CC=C265C266=CC=CC=C266C267=CC=CC=C267C268=CC=CC=C268C269=CC=CC=C269C270=CC=CC=C270C271=CC=CC=C271C272=CC=CC=C272C273=CC=CC=C273C274=CC=CC=C274C275=CC=CC=C275C276=CC=CC=C276C277=CC=CC=C277C278=CC=CC=C278C279=CC=CC=C279C280=CC=CC=C280C281=CC=CC=C281C282=CC=CC=C282C283=CC=CC=C283C284=CC=CC=C284C285=CC=CC=C285C286=CC=CC=C286C287=CC=CC=C287C288=CC=CC=C288C289=CC=CC=C289C290=CC=CC=C290C291=CC=CC=C291C292=CC=CC=C292C293=CC=CC=C293C294=CC=CC=C294C295=CC=CC=C295C296=CC=CC=C296C297=CC=CC=C297C298=CC=CC=C298C299=CC=CC=C299C300=CC=CC=C300C301=CC=CC=C301C302=CC=CC=C302C303=CC=CC=C303C304=CC=CC=C304C305=CC=CC=C305C306=CC=CC=C306C307=CC=CC=C307C308=CC=CC=C308C309=CC=CC=C309C310=CC=CC=C310C311=CC=CC=C311C312=CC=CC=C312C313=CC=CC=C313C314=CC=CC=C314C315=CC=CC=C315C316=CC=CC=C316C317=CC=CC=C317C318=CC=CC=C318C319=CC=CC=C319C320=CC=CC=C320C321=CC=CC=C321C322=CC=CC=C322C323=CC=CC=C323C324=CC=CC=C324C325=CC=CC=C325C326=CC=CC=C326C327=CC=CC=C327C328=CC=CC=C328C329=CC=CC=C329C330=CC=CC=C330C331=CC=CC=C331C332=CC=CC=C332C333=CC=CC=C333C334=CC=CC=C334C335=CC=CC=C335C336=CC=CC=C336C337=CC=CC=C337C338=CC=CC=C338C339=CC=CC=C339C340=CC=CC=C340C341=CC=CC=C341C342=CC=CC=C342C343=CC=CC=C343C344=CC=CC=C344C345=CC=CC=C345C346=CC=CC=C346C347=CC=CC=C347C348=CC=CC=C348C349=CC=CC=C349C350=CC=CC=C350C351=CC=CC=C351C352=CC=CC=C352C353=CC=CC=C353C354=CC=CC=C354C355=CC=CC=C355C356=CC=CC=C356C357
```

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 200376 bases at least Q40
Consensus quality: 210297 bases at least Q30

Consensus quality: 214722 bases at least Q20
Insert size: 234000; agarose-fp

Insert size: 217780; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

- * runs of N , but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence

- * as soon as it is available and the accession number will
- * be preserved.

★	1	2203: contig of 2203 bp in length
★	2204	2303: gap of 100 bp
★		

* 3531: conig of 1228 bp in length
* 3631: con of 100 bn

*
*
*
3632
4978
4977: conig of 1346 bp in length
5077: con of 100 bp

26778	32091:	contig of 5314 bp	in length
32092	32191:	gap of 100 bp	
32192	32216:	contig of 6025 bp	in length
32217	32316:	gap of 100 bp	
32317	44697:	contig of 6381 bp	in length
44698	44797:	gap of 100 bp	
44798	50192:	contig of 5395 bp	in length
50193	50292:	gap of 100 bp	
50293	56609:	contig of 6317 bp	in length
56610	56709:	gap of 100 bp	
56710	65009:	contig of 8300 bp	in length
65010	65109:	gap of 100 bp	
65110	74207:	contig of 9098 bp	in length
74208	74307:	gap of 100 bp	
74308	81141:	contig of 6834 bp	in length
81142	81241:	gap of 100 bp	
81242	99006:	contig of 17765 bp	in length
99007	99106:	gap of 100 bp	
99107	112940:	contig of 13834 bp	in length
112941	113040:	gap of 100 bp	
113041	125171:	contig of 12131 bp	in length
125172	125271:	gap of 100 bp	
125272	136515:	contig of 11244 bp	in length
136516	136615:	gap of 100 bp	
136616	148843:	contig of 12234 bp	in length
148844	148949:	gap of 100 bp	
148950	161417:	contig of 12468 bp	in length
161418	161517:	gap of 100 bp	
161518	176433:	contig of 14916 bp	in length
176434	176533:	gap of 100 bp	
176534	194704:	contig of 18171 bp	in length
194705	194804:	gap of 100 bp	
194805	210058:	contig of 15254 bp	in length
210059	210158:	gap of 100 bp	
210159	220480:	contig of 10322 bp	in length

FEATURES	source
Location/Qualifiers	1. 220480

```

/organism="Homo sapiens"
/mol_type="genomic DNA"

```

```

/mol_type= genomic DNA
/db_xref="taxon:9606"
/chromosome="12"

```

```

/chromosome="12"
/map="12"
/cloc="pp1-20220"

```

```

/clone="RPCI-793A9"
/clone_lib="RPCI-11 Human Male BAC"

```

```

1. .2203
/note="assembly_fragment

```

```
clone_end:SP6
vector_side:left"
cc:

```

```
2304. .3531
      /note="assembly_fragment"
```

```
3632. .4977
/note="assembly_fragment"
```



```

/misc_feature /note="assembly_fragment"
50293. .56609
/note="assembly_fragment"
56710. .65009
/note="assembly_fragment"
65110. .74207
/note="assembly_fragment"
74308. .81141
/note="assembly_fragment"
81242. .99006
/note="assembly_fragment"
99107. .112940
/note="assembly_fragment"
113041. .125171
/note="assembly_fragment"
125272. .136515
/note="assembly_fragment"
136616. .148849
/note="assembly_fragment"
148950. .161417
/note="assembly_fragment"
161518. .176433
/note="assembly_fragment"
176534. .194704
/note="assembly_fragment"
194805. .210058
/note="assembly_fragment"
210159. .220480
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match 46.3%; Score 19; DB 2; Length 220480;
Best Local Similarity 100.0%; Pred.No. 0.42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AACTACATACACACCT 24
|||||
Db 117356 AACTACATACACACCT 117374

RESULT 9
AX655811 2000 bp DNA linear PAT 22-MAR-2003
LOCUS Sequence 5681 from Patent WO03000898.
DEFINITION AX655811
ACCESSION AX655811
VERSION AX655811.1 GI:29158625
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1
AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 5681 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1. .2000
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 43.9%; Score 18; DB 6; Length 2000;
Best Local Similarity 100.0%; Pred.No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TACATACACACCTTA 26
|||||

```

```

Db 758 TACATACACACCTTA 775

RESULT 10
AP005862/c
LOCUS Oryza sativa (japonica cultivar-group) chromosome 9 clone
DEFINITION OSJNBa0038K02, ** SEQUENCING IN PROGRESS **.
ACCESSION AP005862
VERSION AP005862.1 GI:24430392
KEYWORDS HTG; HTGS PHASE2.
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, BAC
clone-OSJNBa0038K02
JOURNAL Published Only in Database (2002)
REFERENCE
2 (bases 1 to 143961)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(B-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
location/Qualifiers
1. .143961
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="OSJNBa0038K02"

ORIGIN
Query Match 43.9%; Score 18; DB 2; Length 143961;
Best Local Similarity 100.0%; Pred.No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 ACCACCTTAGTCAGATAC 35
|||||
Db 112225 ACCACCTTAGTCAGATAC 112208

RESULT 11
CNS08CD7
LOCUS Oryza sativa chromosome 12, . BAC OSJNBb0071117 of library OSJNBb
DEFINITION from chromosome 12 of cultivar Nipponbare of esp. japonica of Oryza
sativa (rice), complete sequence.
ACCESSION AL954853
VERSION AL954853.5 GI:31414507
KEYWORDS HTG.
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 (bases 1 to 159056)

```

AUTHORS
 Choisme N., Orijeda G., Cattolico L., Demange N., Wincker P., Segreane B., Palletier E., Scarpetti C., Salanoubat M., Weissenbach J. and Quetier F.
TITLE
 Oryza sativa chromosome 12 sequencing
REFERENCE
 2 (bases 1 to 159056)
JOURNAL
 Direct Submission
 Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - France (E-mail : segre@genoscope.cns.fr)
 - web : www.genoscope.cns.fr
 On Jun 4, 2003 this sequence version replaced gi:28372080.
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: segre@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : OSJNBa0021D06 (AC-BX000560)
 Downstream BAC (overlapping the SP6 end) : OUL618_C05 (AC-AL713952)
 ----- Finishing boundaries
 FINISHED SEGMENT STARTS AT BASE 5389
 FINISHED SEGMENT ENDS AT BASE 159056

FEATURES
 Location/Qualifiers
 1..159056
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:39947"
 /chromosome="12"
 /clone="OSJNB0071117"
 /clone_lib="OSJNBb"

ORIGIN
 Query Match 43.9%; Score 18; DB 8; Length 159056;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 9 TACATACACACACCTTA 26
 |||||
 Db 108806 TACATACACACACCTTA 108823

RESULT 12
 AC125959
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-169C9, *** SEQUENCING IN PROGRESS
 *** 24 unordered pieces.
 AC125959 215960 bp DNA linear HTG 26-SEP-2002
 AC125959
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 215960)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barbez, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, J., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Maingardine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
 Unpublished
 2 (bases 1 to 215960)
 Worley, K. C.
Direct Submission
 Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 215960)
Rat Genome Sequencing Consortium.
Direct Submission
 Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Sep 26, 2002 this sequence version replaced gi:21671705.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GQYN
 Center clone name: CH230-169C9
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 159310 bases at least Q40
 Consensus quality: 166804 bases at least Q30
 Consensus quality: 170961 bases at least Q20
 Estimated insert size: 200252; sum-of-contigs estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 31293: contig of 31293 bp in length
* 31294 31393: gap of unknown length
* 31394 42669: contig of 11276 bp in length
* 42670 42769: gap of unknown length
* 42770 55798: contig of 13029 bp in length
* 55799 55898: gap of unknown length
* 55899 62523: contig of 6625 bp in length
* 62524 62624: gap of unknown length
* 62624 80012: contig of 17389 bp in length
* 80013 80112: gap of unknown length
* 80113 91361: contig of 11249 bp in length
* 91362 91461: gap of unknown length
* 91462 171227: contig of 79765 bp in length
* 171227 171327: gap of unknown length
* 171327 183652: contig of 12326 bp in length
* 183653 183753: gap of unknown length
* 183753 185305: contig of 1553 bp in length
* 185306 185408: gap of unknown length
* 185409 186445: contig of 1140 bp in length
* 186446 186645: gap of unknown length
* 186646 187846: contig of 1201 bp in length
* 187847 187946: gap of unknown length
* 187947 189065: contig of 1119 bp in length
* 189066 189166: gap of unknown length
* 189167 190763: contig of 1598 bp in length
* 190764 190863: gap of unknown length
* 190864 192694: contig of 1831 bp in length
* 192695 192794: gap of unknown length
* 192795 193943: contig of 1148 bp in length
* 193944 194042: gap of unknown length
* 194043 196133: contig of 2091 bp in length
* 196134 196233: gap of unknown length
* 196234 198533: contig of 2300 bp in length
* 198534 198633: gap of unknown length
* 198634 200628: contig of 1995 bp in length
* 200629 200728: gap of unknown length
* 200729 202216: contig of 1488 bp in length
* 202217 202316: gap of unknown length
* 202317 205020: contig of 2704 bp in length
* 205021 205120: gap of unknown length
* 205121 207746: contig of 2626 bp in length
* 207747 207846: gap of unknown length
* 207847 209300: contig of 1454 bp in length
* 209301 209400: gap of unknown length
* 209401 211460: contig of 2060 bp in length
* 211461 211560: gap of unknown length
* 211561 215960: contig of 4400 bp in length.
  
```

FEATURES

```

Source
1. 215960
   /organism="Rattus norvegicus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10116"
   /clone="CH230-169C9"
misc_feature
21904..24251
   /note="wgs_contig"
misc_feature
31394..33471
   /note="wgs_contig"
misc_feature
55899..57107
   /note="wgs_contig"
misc_feature
80113..82141
   /note="wgs_contig"

```

ORIGIN

```

Query Match      43.9%; Score 18; DB 2; Length 215960;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 AATGAACATACATACAC 19
        |||||
DB      91069 AATGAACATACATACAC 91086

RESULT 13
AC109976      282132 bp      DNA      linear      HTG 26-SEP-2002
LOCUS      Rattus norvegicus clone CH230-177G10, *** SEQUENCING IN PROGRESS
DEFINITION      ***, 25 unordered pieces.
AC109976
VERSION      AC109976.4 GI:23321469
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE
1 (bases 1 to 282132)
AUTHORS      Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Amin,A., Angiano,D.,
              Allen,C., Allen,H., Alsbrooks,S., Ayodeji,M., Baca,E., Baden,H.,
              Ayalabechechi,V., Aoyagi,A., Ayodeji,M., Barnstead,M., Benahmed,F.,
              Baldwin,D., Bandaranaike,D., Barber,M., Blyth,P., Brown,M.,
              Biswal,K., Blair,J., Blankenburg,K., Burch,P., Burrell,K., Calderon,E.,
              Bryant,N., Buhay,C., Burch,P., Burrell,K., Ceasar,H., Center,A.,
              Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,Y., Chen,Z., Chu,J.,
              Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Coyle,M., Cree,A., D'Souza,L.,
              Cleveland,C., Cockrell,R., Cox,C., D'Amico,C., Ding,Y., Dinh,B., Divya,K.,
              Davila,M.L., Davis,C., Davy-Carroll,L., De Andia,C., Dederich,D.,
              Delgado,O., Denison,S., Deramo,C., Evans,C.A., Falls,T., Fan,G.,
              Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K.,
              Egan,A., Escotto,M., Eugene,C., Evans,C.A., Evans,C.A., Falls,T., Fan,G.,
              Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
              Fraser,C.M., Gabisi,A., Garcia,R., Garcia,A., Garner,T., Garza,M.,
              Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Grady,M., Guzman,W., Guevara,W.,
              Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
              Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
              Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
              Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
              Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
              Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
              Kowis,C., Kraft,C.I., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
              Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
              Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
              Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
              Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
              Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
              Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
              Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
              Nwokeneme,O., Okwuonu,G., Olarnunpagoon,A., Pal,S., Parks,K.,
              Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
              Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
              Puazo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R.,
              Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
              Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
              Sanders,W., Savery,G., Schefer,S., Scott,G., Shatsman,S., Shen,H.,
              Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
              Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
              Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
              Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
              Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
              Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
              Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
              Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
              Yu,F., Zhang,J., Zhou,J., Zhou,S., Zhou,S., Dunn,D., von
              Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
              Weinstein,G. and Gibbs,R.A.
              Direct Submission

```

TITLE

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 282132)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Submitted (09-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 282132)

AUTHORS

Rat Genome Sequencing Consortium.

TITLE

Direct Submission

JOURNAL

Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Sep 26, 2002 this sequence version replaced gi:21738290. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GQUN

Center clone name: CH230-177G10

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 215218 bases at least Q40

Consensus quality: 222117 bases at least Q30

Consensus quality: 226410 bases at least Q20

Estimated insert size: 263059; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 64301: contig of 64301 bp in length
 * 64302 64401: gap of unknown length
 * 64402 79310: contig of 14909 bp in length
 * 79311 79410: gap of unknown length
 * 79411 90217: contig of 10807 bp in length
 * 90218 90317: gap of unknown length
 * 90318 105370: contig of 15053 bp in length
 * 105371 105470: gap of unknown length
 * 105471 108531: contig of 3061 bp in length
 * 108532 108631: gap of unknown length
 * 108632 127118: contig of 18487 bp in length
 * 127119 127218: gap of unknown length
 * 127219 153038: contig of 25820 bp in length
 * 153039 153138: gap of unknown length
 * 153139 182729: contig of 29591 bp in length
 * 182730 182829: gap of unknown length
 * 182830 193903: contig of 11074 bp in length
 * 193904 194003: gap of unknown length
 * 194004 206429: contig of 12426 bp in length
 * 206430 206529: gap of unknown length
 * 206530 207753: contig of 1224 bp in length
 * 207754 207853: gap of unknown length
 * 207854 209421: contig of 1568 bp in length
 * 209422 209521: gap of unknown length
 * 209522 211065: contig of 1544 bp in length

* 211066 211165: gap of unknown length
 * 211166 212868: contig of 1703 bp in length
 * 212869 212969: gap of unknown length
 * 212970 216482: contig of 3514 bp in length
 * 216483 216582: gap of unknown length
 * 216583 218604: contig of 2022 bp in length
 * 218605 218704: gap of unknown length
 * 218705 221765: contig of 3061 bp in length
 * 221766 221865: gap of unknown length
 * 221866 225831: contig of 3966 bp in length
 * 225832 225932: gap of unknown length
 * 225933 233131: contig of 7200 bp in length
 * 233132 233231: gap of unknown length
 * 233232 244973: contig of 11742 bp in length
 * 244974 245073: gap of unknown length
 * 245074 246570: contig of 1497 bp in length
 * 246571 246670: gap of unknown length
 * 246671 252671: contig of 6000 bp in length
 * 252672 252770: gap of unknown length
 * 252771 269281: contig of 16511 bp in length
 * 269282 269381: gap of unknown length
 * 269382 275653: contig of 6272 bp in length
 * 275654 275754: gap of unknown length
 * 275755 282132: contig of 6379 bp in length.

FEATURES

source

Location/Qualifiers

1..282132

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-177G10"

43744..45689

/note="wgs contig"

64402..68669

/note="wgs contig"

79411..83544

/note="wgs contig"

105471..106633

/note="wgs contig"

107196..108531

/note="wgs contig"

108632..110536

/note="wgs contig"

113702..115613

/note="wgs contig"

115664..116916

/note="wgs contig"

182830..184137

/note="wgs contig"

191871..193903

/note="wgs contig"

ORIGIN

Query Match 43.9%; Score 18; DB 2; Length 282132;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAACATACATACAAAC 19

Db 258584 AATGAACATACATACAAAC 258601

RESULT 14

AK073122/c

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) cDNA clone:J033003A15, full

insert sequence.

ACCESSION

AK073122

VERSION

AK073122.1 GI:32983145

KEYWORDS

FLI CDNA; CAP trapper.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

AK073122 1712 bp mRNA linear PLN 24-JUL-2003

Ehrhartoideae; Oryzae; Oryza.

REFERENCE AUTHORS

The Rice Full-length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-length cDNA Project Team,
Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K.,
Kishimoto N., Yasaki K., Ishikawa M., Yamada H., Ooka H., Hotta I.,
Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C.,
Ohtsuka K., Shishiki T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ohtsuka Y., Murakami K.,
Iida Y., Sugano S., Fujimura T., Suzuki Y., Tanoda Y.,
Kodama T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M.,
Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Niikura J.,
Ikeda R., Ishibiki J., Kawamata M., Yoshimura A., Miura J.,
Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., KIKEN:
Kawai J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S.,
Hara A., Hashidume W., Hayatsu N., Imotani K., Ishii Y., Itoh M.,
Kagawa I., Kondo S., Konno H., Miyazaki A., Oota N., Oca Y.,
Saito R., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T.,
Yoshino M. and Hayashizaki Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from

TITLE

japonica rice
Science 301 (5631), 376-379 (2003)

JOURNAL

22752273

PUBMED

12869764

REFERENCE AUTHORS

2 (bases 1 to 1712)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishii, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuka, K., Oka, M., Ooka, H.,
Oota, N., Oca, Y., Ohtsuka, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yasaki, J., Yokomizo, S. and
Yoshimura, A.

Direct Submission

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28k full-length cDNA clones from japonica
rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-length cDNA Project Team: Kikuchi S., Satoh K.,
Negata, T., Kawagashira N., Doi K., Kishimoto N., Yasaki J.,
Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T.,
Ohneda E., Yahagi W., Suzuki K., Li C., Ohtsuka K., Shishiki T. and
Yamamoto M.

FAIS Genome Sequencing & Analysis Group: Ohtsuka Y., Iida Y.,
Fujimura T., Ikeda R., Ishibiki J., Kawamata M., Kobayashi M.,
Kodama T., Kurotsaki T., Kusumegi T., Lu M., Masuda H., Miura J.,
Mizuno K., Narikawa R., Niikura J., Oka M., Ryu R., Sugano S.,
Sugiyama A., Suzuki Y., Tanoda Y., Ueda M., Xie Q., Yokomizo S.,
Yoshimura A., Matsubara K. and Murakami K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi J., Aizawa K.,
Akimura T., Arakawa T., Carninci P., Fukuda S., Hanagaki T.,
Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K.,
Hiraoka T., Hori F., Iida J., Imamura K., Imotani K., Ishii Y.,
Itoh M., Kagawa I., Kanagawa S., Katoh H., Kawai J.,
Kishikawa-Hirozane, T., Kojima Y., Kondo S., Konno H., Kouda M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Oota, N.,
Oca, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES

source

1. 1712
/location=Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033003A15"

ORIGIN

Query Match 41.5%; Score 17; DB 8; Length 1712;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 ACCACCTTAGTCAGATA 34

Db 646 ACCACCTTAGTCAGATA 630

RESULT 15

AX346519/c

LOCUS AX346519 16918 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 1590 from Patent WO0200928.
ACCESSION AX346519
VERSION AX346519.1 GI:18494405

KEYWORDS

synthetic construct
synthetic construct
artificial sequences.

SOURCE

ORGANISM

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1590 03-JAN-2002;
EpiGenomics AG (DE)

FEATURES

source

1. 16918
/location=Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 41.5%; Score 17; DB 6; Length 16918;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ACTACATAACACCCACC 23

Db 6093 ACTACATAACACCCACC 6077

Search completed: March 25, 2004, 02:32:49
Job time : 1708 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 00:14:02 ; Search time 249 Seconds
(without alignments)
699,502 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42
Perfect score: 41
Sequence: 1 gaatgaactacatacaaac.....ccttagtcagatcactctt 41

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 337863 seqs, 212409041 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq_29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1431	3	AAA71435 Human meg
2	40	97.6	128	3	AAA71434 Human meg
3	30	73.2	30	3	AAA71449 Human meg
4	30	73.2	30	3	AAA71448 Human meg
5	25	61.0	25	3	AAA71453 Human meg
6	23	56.1	23	3	AAA71452 Human meg
7	21	51.2	6182	6	ABL34015 Human inn
8	19	46.3	1523	3	ACAC39949 Arabidops
9	18	43.9	2000	7	ADA72356 Rice gene
10	17	41.5	16918	6	ABL33617 Human inn
11	17	41.5	50000	8	ADB16926 Human DYX
12	16	39.0	1652	4	ABL07443 Drosophil
13	16	39.0	3991	6	ABL07442 Drosophil
14	16	39.0	6160	6	ABK31272 Signal tr
15	16	39.0	6160	6	ABL70233 Chemical
16	16	39.0	17993	6	AAD46721 Human tra
17	15	36.6	241	3	AAC13384 Human sec
18	15	36.6	756	4	AAH71525 Human cer
19	15	36.6	789	7	ACA44540 Prokaryot
20	15	36.6	951	7	ACF72327 Staphyloc
21	15	36.6	1002	2	AAT13163 SCSV segm
22	15	36.6	1188	5	AAS89552 DNA encod
23	15	36.6	1189	5	AAS87858 DNA encod

24	15	36.6	1254	5	AAS89553 DNA encod
25	15	36.6	1617	4	AAS53391 Haemophil
26	15	36.6	1617	7	ACA34260 Prokaryot
27	15	36.6	1698	4	AAH73359 Human cer
28	15	36.6	1698	5	ABV25202 Human pro
29	15	36.6	1698	5	ABV24391 Human pro
30	15	36.6	1886	3	AAC69532 Human sec
31	15	36.6	1886	7	ADA55988 Gene enco
32	15	36.6	1886	7	ADA39801 Human sec
33	15	36.6	1886	7	ACC50434 Human sec
34	15	36.6	1966	6	ABI99599 Mouse isc
35	15	36.6	1986	2	AAQ97797 Clone prd
36	15	36.6	2084	4	ABL10239 Drosophil
37	15	36.6	2360	9	ADB63696 Human cDN
38	15	36.6	5926	6	ABL33653 Human inn
39	15	36.6	6049	6	ABL32228 Human inn
40	15	36.6	6052	4	AAS59391 Mouse cDN
41	15	36.6	6075	4	AAS59389 Mouse cDN
42	15	36.6	6120	4	AAS59390 Mouse cDN
43	15	36.6	6191	6	ABL33216 Human inn
44	15	36.6	6191	6	ABK31306 Signal tr
45	15	36.6	6191	6	ABL70281 Chemical

ALIGNMENTS

RESULT 1
AAA71435
ID AAA71435 standard; DNA; 1431 BP.
XX
AC AAA71435;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter fragment DNA.
XX
KW Promoter; megsin; human; protein isolation; screening. ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megsin useful for screening proteins.
XX
PS Disclosure; Fig 2; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents a fragment of the human megsin promoter which is described in the method of the invention

SQ Sequence 1431 BP; 466 A; 266 C; 303 G; 394 T; 0 U; 2 Other;
Query Match 100.0%; Score 41; DB 3; Length 1431;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATGAACCTACATACCAACCACTTGTAGTCAGATACTACTTT 41
Db 1303 GAATGAACCTACATACCAACCACTTGTAGTCAGATACTACTTT 1343

RESULT 2
AAA71434
ID AAA71434 standard; DNA; 128 BP.
XX AC AAA71434;
XX 01-DEC-2000 (first entry)
XX Human megasin promoter fragment DNA.
XX Promoter; megasin; human; protein isolation; screening. ss.
XX Homo sapiens.
XX W0200043528-A1.
XX 27-JUL-2000.
XX 25-JAN-2000; 2000WO-JP000350.
XX 25-JAN-1999; 99JP-00015667.
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX Miyata T;
XX WPI; 2000-543257/49.
XX DNA for promoter region of megasin useful for screening proteins.
XX Example 5; Page 40; 45pp; Japanese.
XX This invention describes a novel DNA sequence (I) representing a promoter
region having part or all of a specific base sequence. The invention also
describes (1) a vector containing (I); (2) a cell transformed by the
above vector; and (3) protein produced using (I). (I) is useful for
screening and isolating proteins (especially transcription factors).
XX AAA71434-A71469 represent PCR primers used in the method described in the
invention
XX Sequence 30 BP; 12 A; 9 C; 2 G; 7 T; 0 U; 0 Other;
XX Query Match 73.2%; Score 30; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 TACATAACCAACCACTTGTAGTCAGATACTAC 38
Db 1 TACATAACCAACCACTTGTAGTCAGATACTAC 30

RESULT 4
AAA71448
ID AAA71448 standard; DNA; 30 BP.
XX AC AAA71448;
XX 01-DEC-2000 (first entry)
XX Human megasin promoter PCR primer SEQ ID NO: 15.
XX Promoter; megasin; human; protein isolation; screening. PCR primer; ss.
XX Homo sapiens.
XX W0200043528-A1.
XX 27-JUL-2000.
XX 25-JAN-2000; 2000WO-JP000350.
XX 25-JAN-1999; 99JP-00015667.
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX Miyata T;
XX WPI; 2000-543257/49.
XX DNA for promoter region of megasin useful for screening proteins.
XX Example 5; Page 40; 45pp; Japanese.
XX This invention describes a novel DNA sequence (I) representing a promoter
region having part or all of a specific base sequence. The invention also
describes (1) a vector containing (I); (2) a cell transformed by the
above vector; and (3) protein produced using (I). (I) is useful for
screening and isolating proteins (especially transcription factors). This
sequence represents the human megasin promoter which is described in the
method of the invention
XX Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
XX Query Match 97.6%; Score 40; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 5.2e-13;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AATGAACCTACATACCAACCACTTGTAGTCAGATACTACTTT 41
Db 1 AATGAACCTACATACCAACCACTTGTAGTCAGATACTACTTT 40

RESULT 3
AAA71449
ID AAA71449 standard; DNA; 30 BP.
XX AC AAA71449;
XX 01-DEC-2000 (first entry)
XX Human megasin promoter PCR primer SEQ ID NO: 16.
XX Promoter; megasin; human; protein isolation; screening. PCR primer; ss.
XX Homo sapiens.
XX W0200043528-A1.
XX
```

XX
AC AAA71452:XX
BE
03-TIT-2001-2001WC-EP007537


```

XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PA (EPIC-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
XX PT diagnosis and treatment of diseases associated with abnormal cytosine
XX PT methylation.
XX PS Claim 1; SEQ ID NO 1989; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention
XX SQ Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;

Query Match 51.2%; Score 21; DB 6; Length 6182;
Best Local Similarity 100.0%; Pred. No. 0.029;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AACTACATACACACCTTA 26
Db 1306 AACTACATACACACCTTA 1286

RESULT 8
AAC39949/c
ID AAC39949 standard; DNA; 1523 BP.
XX AC AAC39949;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26493.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway; metabolic pathway;
XX KW promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 29-MAR-1999; 99US-0126264P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129845P.
XX PR 19-APR-1999; 99US-0130077P.
XX PR 21-APR-1999; 99US-0130449P.
XX PR 23-APR-1999; 99US-0130510P.
XX PR 23-APR-1999; 99US-0130891P.
XX PR 28-APR-1999; 99US-0131449P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 04-MAY-1999; 99US-0132407P.
XX PR 05-MAY-1999; 99US-0132484P.
XX PR 06-MAY-1999; 99US-0132485P.
XX PR 06-MAY-1999; 99US-0132486P.
XX PR 07-MAY-1999; 99US-0132487P.
XX PR 11-MAY-1999; 99US-0132863P.
XX PR 14-MAY-1999; 99US-0134218P.
XX PR 14-MAY-1999; 99US-0134219P.
XX PR 14-MAY-1999; 99US-0134221P.
XX PR 14-MAY-1999; 99US-0134370P.
XX PR 18-MAY-1999; 99US-0134370P.
XX PR 19-MAY-1999; 99US-0134768P.
XX PR 20-MAY-1999; 99US-0134941P.
XX PR 21-MAY-1999; 99US-0135124P.
XX PR 21-MAY-1999; 99US-0135353P.
XX PR 24-MAY-1999; 99US-0135629P.
XX PR 25-MAY-1999; 99US-0136021P.
XX PR 27-MAY-1999; 99US-0136392P.
XX PR 28-MAY-1999; 99US-0136782P.
XX PR 01-JUN-1999; 99US-0137222P.
XX PR 03-JUN-1999; 99US-0137528P.
XX PR 04-JUN-1999; 99US-0137502P.
XX PR 07-JUN-1999; 99US-0137724P.
XX PR 08-JUN-1999; 99US-0138094P.
XX PR 10-JUN-1999; 99US-0138540P.
XX PR 10-JUN-1999; 99US-0138847P.
XX PR 14-JUN-1999; 99US-0139119P.
XX PR 16-JUN-1999; 99US-0139452P.
XX PR 16-JUN-1999; 99US-0139453P.
XX PR 17-JUN-1999; 99US-0139492P.
XX PR 18-JUN-1999; 99US-0139454P.
XX PR 18-JUN-1999; 99US-0139455P.
XX PR 18-JUN-1999; 99US-0139456P.
XX PR 18-JUN-1999; 99US-0139457P.
XX PR 18-JUN-1999; 99US-0139458P.
XX PR 18-JUN-1999; 99US-0139459P.
XX PR 18-JUN-1999; 99US-0139460P.
XX PR 18-JUN-1999; 99US-0139461P.
XX PR 18-JUN-1999; 99US-0139462P.
XX PR 18-JUN-1999; 99US-0139463P.
XX PR 18-JUN-1999; 99US-0139750P.
XX PR 18-JUN-1999; 99US-0139763P.
XX PR 21-JUN-1999; 99US-0139817P.
XX PR 21-JUN-1999; 99US-0139899P.
XX PR 23-JUN-1999; 99US-0140353P.
XX PR 23-JUN-1999; 99US-0140354P.
XX PR 24-JUN-1999; 99US-0140695P.
XX PR 28-JUN-1999; 99US-0140823P.
XX PR 28-JUN-1999; 99US-0140991P.
XX PR 30-JUN-1999; 99US-0141287P.
XX PR 01-JUL-1999; 99US-0141842P.
XX PR 02-JUL-1999; 99US-0142154P.
XX PR 02-JUL-1999; 99US-0142055P.
XX PR 06-JUL-1999; 99US-0142390P.
XX PR 08-JUL-1999; 99US-0142803P.
XX PR 09-JUL-1999; 99US-0142920P.
XX PR 12-JUL-1999; 99US-0142977P.
XX PR 13-JUL-1999; 99US-0143542P.
XX PR 14-JUL-1999; 99US-0143624P.
XX PR 15-JUL-1999; 99US-0144005P.
XX PR 16-JUL-1999; 99US-0144085P.
XX PR 16-JUL-1999; 99US-0144086P.
XX PR 19-JUL-1999; 99US-0144325P.
XX PR 19-JUL-1999; 99US-0144331P.
XX PR 19-JUL-1999; 99US-0144332P.
XX PR 19-JUL-1999; 99US-0144333P.
XX PR 19-JUL-1999; 99US-0144334P.
XX PR 19-JUL-1999; 99US-0144335P.
XX PR 20-JUL-1999; 99US-0144352P.
XX PR 20-JUL-1999; 99US-0144632P.
XX PR 20-JUL-1999; 99US-0144884P.
XX PR 21-JUL-1999; 99US-0144814P.

```

```
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-015984P.

PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 46.3%; Score 19; DB 3; Length 1523;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 CAACCACTTAGTCAGATA 34
Db 646 CAACCACTTAGTCAGATA 628

RESULT 9
ADA72356
ID ADA72356 standard; DNA; 2000 BP.
XX AC ADA72356;
XX DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 5681.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PA Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX PS Claim 27; SEQ ID NO 5681; 899pp; English.
XX CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
```

CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.

XX
SQ Sequence 2000 BP; 576 A; 406 C; 423 G; 593 T; 0 U; 2 Other;

Query Match 43.9%; Score 18; DB 7; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TACATAACACACCTTA 26
|||||
Db 758 TACATAACACACCTTA 775

RESULT 10
ABL33617/C
ID ABL33617 standard; DNA; 16918 BP.
XX
AC ABL33617;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1590.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.

XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1590; 32pp + Sequence Listing; German.

XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention

XX
SQ Sequence 16918 BP; 4516 A; 235 C; 3886 G; 8281 T; 0 U; 0 Other;

Query Match 41.5%; Score 17; DB 6; Length 16918;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ACTACATAACACACCTT 23

Db 6093 ACTACATAACACACCTT 6077
|||||

RESULT 11
ADB16926
ID ADB16926 standard; DNA; 50000 BP.
XX
AC ADB16926;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human DYXC1 DNA, chromosomal gene region nucleotides 1-50000.
XX
KW ds; human; DYXC1; dyslexia; neurological disorder; chromosome 15q21;
KW reading disability; phonological processing; rapid naming;
KW verbal short-term memory.
XX
XX Homo sapiens.
XX
XX WO2003068814-A1.
XX
XX 21-AUG-2003.
XX
XX 12-FEB-2003; 2003WO-FI000110.
XX
XX 12-FEB-2002; 2002US-0355782P.
XX
XX (LICN) LICENTIA LTD.
XX
XX Kere J, Taipale M, Nopola-Hemmi J, Kaminen N;
XX
XX WPI; 2003-646482/61.
XX
XX New isolated, purified DYXC1 nucleic acid for studying brain processes,
XX e.g. reading, phonological processing, rapid naming or verbal short-term
XX memory, or for diagnosing dyslexia or assessing the predisposition to
XX dyslexia.

XX Claim 27; Page 55-69; 135pp; English.

XX
XX This invention relates to a novel isolated human gene DYXC1 that is
XX functionally related to dyslexia, more particularly it describes single
XX nucleotide polymorphisms thought to predispose an individual in to
XX developing dyslexia. This is a neurological disorder with a genetic basis
XX (DYXC1 has been isolated to chromosome 15q21), which manifests itself as
XX a specific reading disability. Specifically, DYXC1 is can be useful in
XX study of brain processes such as reading, phonological processing, rapid
XX naming and verbal short-term memory. Accordingly, the present invention
XX describes methods and materials for analysing allelic variations in the
XX DYXC1 gene, and also provides DYXC1 as an antigen for the production of
XX antibodies used in the diagnosis of dyslexia. This polynucleotide is the
XX partial genomic sequence of the human DYXC1 chromosomal region
XX (nucleotides 1-50000) of the invention.

XX
SQ Sequence 50000 BP; 15686 A; 9567 C; 10158 G; 14589 T; 0 U; 0 Other;

Query Match 41.5%; Score 17; DB 8; Length 50000;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CTACATAACACACCTT 24
|||||

Db 24331 CTACATAACACACCTT 24347

RESULT 12
ABL07443
ID ABL07443 standard; cDNA; 1652 BP.
XX
AC ABL07443;
XX
DT 26-MAR-2002 (first entry)

```

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16811.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX FI WPI; 2001-656860/75.
XX DR P-PSDB; AB963340.
XX XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 16811; 21pp + Sequence Listing; English.
XX PS The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1652 BP; 434 A; 391 C; 406 G; 421 T; 0 U; 0 Other;
XX
XX Query Match 39.0%; Score 16; DB 4; Length 1652;
XX Best Local Similarity 100.0%; Pred. No. 20;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 16 CAACCACTTACTGTCAG 31
XX Db |||||
XX
XX RESULT 13
XX ABL07442/C
XX ID ABL07442 standard; cDNA; 3991 BP.
XX AC
XX XX
XX XX 26-MAR-2002 (first entry)
XX DT
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16808.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR
XX PT

```

```

PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX FI WPI; 2001-656860/75.
XX DR P-PSDB; ABB63339.
XX XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 16808; 21pp + Sequence Listing; English.
XX PS The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 3991 BP; 1159 A; 901 C; 865 G; 1066 T; 0 U; 0 Other;
XX
XX Query Match 39.0%; Score 16; DB 4; Length 3991;
XX Best Local Similarity 100.0%; Pred. No. 20;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 16 CAACCACTTACTGTCAG 31
XX Db |||||
XX
XX RESULT 14
XX ABR31272/C
XX ID ABR31272 standard; DNA; 6160 BP.
XX AC
XX XX
XX XX 23-APR-2002 (first entry)
XX DT
XX DE Signal transduction associated gene modified DNA #58.
XX KW Human; signal transduction associated gene; cytosine methylation state;
XX KW CpG island; signal transduction associated disease; solid tumour; cancer;
XX KW antitumour; cytostatic; mutant; ds.
XX OS Homo sapiens.
XX OS Synthetic.
XX XX
XX XX WO200200926-A2.
XX PN
XX PD 03-JAN-2002.
XX DE
XX PF 29-JUN-2001; 2001WO-EP007472.
XX XX
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX XX
XX PA (EPIG-) EPIGENOMICS AG.
XX XX
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-147896/19.
XX XX
XX PT Oligonucleotide for diagnosis and therapy of diseases associated with
XX PT signal transduction e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with signal transduction.

```

XX Claim 1; SEQ ID NO 115; 24pp; English.
 XX The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphate or disulphite.
 CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
 CC the cytosine methylation state (CpG islands) of these genes, and a method
 CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
 CC genes associated with signal transduction. The genomic DNA can be
 CC obtained from cells or cellular components which contain DNA, e.g. cell
 CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
 CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
 CC brain, heart, prostate, lung, breast or liver, histologic object slides,
 CC and all their possible combinations. The sequences of the invention are
 CC useful for the diagnosis and therapy of diseases associated with signal
 CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
 CC chemically pretreated genomic DNA sequences of different genes associated
 CC with signal transduction, or their complementary sequences. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office
 XX
 SQ Sequence 6160 BP; 1580 A; 139 C; 1637 G; 2804 T; 0 U; 0 Other;

Query Match 39.0%; Score 16; DB 6; Length 6160;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACTACATACACACCAC 22
 Db 68 ACTACATACACACCAC 53

RESULT 15
 ABL70233/c
 ID ABL70233 standard; DNA; 6160 BP.
 AC ABL70233;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated cell signalling DNA sequence#62.
 XX
 KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
 XX tumour; cytostatic; ds.
 XX
 OS Unidentified.
 XX
 EN WO200202807-A2.
 XX
 PD 10-JAN-2002.
 XX
 XX 29-JUN-2001; 2001WO-EP007471.
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-154758/20.
 DR
 XX
 XX Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signalling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signalling.
 XX
 XX Claim 1; SEQ ID NO 123; 24pp + Sequence Listing; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signalling. The activity of the modified sequences of the

CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling. Note: The
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office
 XX

SQ Sequence 6160 BP; 1580 A; 139 C; 1637 G; 2804 T; 0 U; 0 Other;
 Query Match 39.0%; Score 16; DB 6; Length 6160;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACTACATACACACCAC 22
 Db 68 ACTACATACACACCAC 53

Search completed: March 25, 2004, 02:04:13
 Job time : 252 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 01:51:12 ; Search time 2010 Seconds
(without alignments)
609.129 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42

Perfect score: 41
Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactactttt 41

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
C 1	19	46.3	294 10	BF365378 QV4-NT002
C 2	18	43.9	549 9	AI188862 qd27a12.x
C 3	18	43.9	931 29	CNS02BQG
C 4	17	41.5	72 14	CF608248 GEMMA01_0

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C	5	17	41.5	228	9	AV319739
C	6	17	41.5	408	28	AV319739
C	7	17	41.5	484	28	AV319739
C	8	17	41.5	531	28	AV319739
C	9	17	41.5	567	28	AV319739
C	10	17	41.5	569	28	AV319739
C	11	17	41.5	649	14	AV319739
C	12	17	41.5	667	13	AV319739
C	13	17	41.5	695	13	AV319739
C	14	17	41.5	782	29	AV319739
C	15	17	41.5	870	13	AV319739
C	16	17	41.5	927	28	AV319739
C	17	17	41.5	1036	10	AV319739
C	18	17	41.5	3764	11	AV319739
C	19	16	39.0	180	14	AV319739
C	20	16	39.0	180	14	AV319739
C	21	16	39.0	259	10	AV319739
C	22	16	39.0	300	9	AV319739
C	23	16	39.0	381	10	AV319739
C	24	16	39.0	397	14	AV319739
C	25	16	39.0	402	9	AV319739
C	26	16	39.0	416	14	AV319739
C	27	16	39.0	427	13	AV319739
C	28	16	39.0	436	10	AV319739
C	29	16	39.0	438	28	AV319739
C	30	16	39.0	460	29	AV319739
C	31	16	39.0	484	10	AV319739
C	32	16	39.0	485	10	AV319739
C	33	16	39.0	486	14	AV319739
C	34	16	39.0	486	14	AV319739
C	35	16	39.0	513	14	AV319739
C	36	16	39.0	517	10	AV319739
C	37	16	39.0	528	12	AV319739
C	38	16	39.0	533	14	AV319739
C	39	16	39.0	535	12	AV319739
C	40	16	39.0	544	28	AV319739
C	41	16	39.0	563	28	AV319739
C	42	16	39.0	564	10	AV319739
C	43	16	39.0	569	9	AV319739
C	44	16	39.0	574	12	AV319739
C	45	16	39.0	576	14	AV319739

ALIGNMENTS

RESULT 1
BF365378/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BF365378 294 bp mRNA linear EST 24-NOV-2000
QV4-NT0028-080700-286-c08 NT0028 Homo sapiens cDNA, mRNA sequence.

BF365378.1 GI:11327403

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 294)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,

O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: aaimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&t2=QV4-NT0028-080700-286-C08&t3=2000-07-08&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 294.

FEATURES

source

1. .294
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NT0028"
/note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 46.3%; Score 19; DB 10; Length 294;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AACTACATACACCACT 24

|||||

Db 112 AACTACATACACCACT 94

RESULT 2

A1188862/c

LOCUS

DEFINITION qd27a12.x1 Soares_placenta 8to9weeks 2NbHPto9W Homo sapiens cDNA
IMAGE:1724926 3' similar to gb:X15414 ALDOSE REDUCTASE
(HUMAN); mRNA sequence.

ACCESSION A1188862

VERSION A1188862.1 GI:3740071

KEYWORDS EST.

SOURCE Homo sapiens

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

(bases 1 to 549)

NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: ccapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality

Insert Length: 1099 Std Error: 0.00
Seq primer: -40UP from Gibco

High quality sequence stop: 1.
Location/Qualifiers

FEATURES

source

1. .549
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1724926"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_placenta 8to9weeks 2NbHPto9W"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a

ORIGIN

Query Match 43.9%; Score 18; DB 9; Length 549;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 ACCACCTTAGTCAGATAC 35

|||||

Db 461 ACCACCTTAGTCAGATAC 444

RESULT 3

CNS02BQG

LOCUS

DEFINITION CNS02BQG 931 bp DNA linear GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
253B17 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL189827

VERSION AL189827.1 GI:7827931

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE 1

AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W., and Weissenbach, J.

Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence

JOURNAL Nat. Genet. 25 (2), 235-238 (2000)

MEDLINE 20296633

PubMed 10835645

REFERENCE 2

AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A., and Weissenbach, J.

Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Genome Res. 10 (7), 939-949 (2000)

MEDLINE 20359837

PubMed 10899143

REFERENCE 3

AUTHORS

Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequençage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr)

JOURNAL

MEDLINE

PubMed

REFERENCE

COMMENT

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES

source

1. .931
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="253B17"
/clone_lib="G"
/notes="Genoscope sequence ID : COAG253CA09SP1-end ;
PUC-Ori"

ORIGIN


```

QY 1 GAATGAACATACATAACA 17
Db 182 GAATGAACATACATAACA 166

RESULT 6
AQ983222
LOCUS
DEFINITION
  RPCI-23-307F19-TV RPCI-23 Mus musculus genomic clone
  RPCI-23-307F19, genomic survey sequence.
ACCESSION
  AQ983222
VERSION
  AQ983222.1 GI:6816427
KEYWORDS
  GSS.
ORGANISM
  Mus musculus (house mouse)
REFERENCE
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
  1 (bases 1 to 408)
AUTHORS
  Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S.,
  Akinret, B., Levins, M., McGann, S., Teegaye, G., Geer, K., Krol, M., de
  Jong, P. and Fraser, C.M.
  Mouse BAC End Sequences from Library RPCI-23
  Other GSSs: RPCI-23-307F19.TJ
  Unpublished (1999)
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-23. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.htm)
  or from Resea ch Genetics (info@resgen.com). BAC end page:
  http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
  Plate: 307 row: F column: 19
  Seq primer: T7
  Class: BAC ends.

FEATURES
  source
  1..408
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="RPCI-23-307F19"
  /sex="Female"
  /lab_host="DH10B"
  /clone_lib="RPCI-23"
  /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
  EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
  brain genomic DNA was isolated and partially digested
  with a combination of EcoRI and EcoRI Methylase. Size
  selected DNA was cloned into the pBACe3.6 vector at the
  EcoRI sites. The ligation products were transformed into
  DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
  Query Match 41.5%; Score 17; DB 28; Length 408;
  Best Local Similarity 100.0%; Pred. No. 72;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CTACATAACACCACT 24
Db 381 CTACATAACACCACT 397

RESULT 7
AQ471300/c
LOCUS
DEFINITION
  AQ471300
  CITBI-E1-2584D24-TR CITBI-E1 Homo sapiens genomic clone 2584D24,
  genomic survey sequence.
ACCESSION
  AQ471300
VERSION
  AQ471300.1 GI:4654954
KEYWORDS
  GSS.
ORGANISM
  Homo sapiens (human)
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 484)
AUTHORS
  Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and
  Venter, J.C.
  Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
  Map Building
  Unpublished (1997)
  Other GSSs: CITBI-E1-2584D24.TF
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbe@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  Seq primer: M13 Reverse
  Class: BAC ends.

FEATURES
  source
  1..484
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /clone="2584D24"
  /sex="male"
  /cell_type="sperm"
  /clone_lib="CITBI-E1"
  /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
  Caltech Human BAC Library D"

ORIGIN
  Query Match 41.5%; Score 17; DB 28; Length 484;
  Best Local Similarity 100.0%; Pred. No. 74;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CTACATAACACCACT 24
Db 332 CTACATAACACCACT 316

RESULT 8
AQ807964
LOCUS
DEFINITION
  AQ807964
  HS 3155_A2_F03_T7C CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=3155 Col=6 Row=K, genomic survey
  sequence.
ACCESSION
  AQ807964
VERSION
  AQ807964.1 GI:5727206
KEYWORDS
  GSS.
ORGANISM
  Homo sapiens (human)
REFERENCE
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 531)
AUTHORS
  Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
  Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
  Hood, L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  99380589
  10449764
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington

QY 8 CTACATAACACCACT 24
Db 381 CTACATAACACCACT 397

RESULT 8
AQ807964
LOCUS
DEFINITION
  AQ807964
  HS 3155_A2_F03_T7C CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=3155 Col=6 Row=K, genomic survey
  sequence.
ACCESSION
  AQ807964
VERSION
  AQ807964.1 GI:5727206
KEYWORDS
  GSS.
ORGANISM
  Homo sapiens (human)
REFERENCE
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 531)
AUTHORS
  Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
  Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
  Hood, L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  99380589
  10449764
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington

```

401 Queen Anna Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@reagen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3155 row: K column: 6
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 531.

FEATURES

source
 1. 531
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clones="Plate=3155 Col=6 Row=K"
 /sex="male"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 41.5%; Score 17; DB 28; Length 531;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CTACATAACCAACCACT 24
 |||||
 Db 460 CTACATAACCAACCACT 476

RESULT 9

AZ369484/c
 LOCUS
 DEFINITION
 1M0120P05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0120P05 F, genomic survey sequence.

ACCESSION
 AZ369484
 VERSION
 AZ369484.1 GI:10483184
 KEYWORDS
 GSS.

SOURCE
 Mus musculus (house mouse)

ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 567)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A., and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

TITLE

Unpublished (2000)

JOURNAL

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0120 row: P column: 05

Seq primer: CGTGTAAACGACGGCCAGT

Class: plasmid ends

High quality sequence stop: 567.

Location/Qualifiers

1. 567

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0120P05"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

FEATURES

source

/clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 Kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 41.5%; Score 17; DB 28; Length 567;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 TTAGTCAGATCTACTT 40
 |||||
 Db 384 TTAGTCAGATCTACTT 368

RESULT 10

CC162058/c
 LOCUS
 DEFINITION
 CC162058
 CC162058
 CC162058.1 GI:30188242
 KEYWORDS
 GSS.

SOURCE

Zea mays
 Zea mays
 Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 569)
 Rabinowicz, P.D., O'Shaughnessy, A.L., Ballia, V., Dedhia, N.,
 Katzenburger, F., King, L., Miller, B., Muller, S., Nascentino, L.,
 Zlatav, T., McCombie, W.R., and Martienssen, R.A.,
 Genomic shotgun sequences from Zea mays (methyl-filtered)
 Unpublished (2002)

ORGANISM

Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mcombie@cshl.org
 Plate: i181 row: b column: 04
 Seq primer: -21M13UnivFwd
 Class: shotgun
 High quality sequence stop: 569.

REFERENCE

1 (bases 1 to 569)
 Rabinowicz, P.D., O'Shaughnessy, A.L., Ballia, V., Dedhia, N.,
 Katzenburger, F., King, L., Miller, B., Muller, S., Nascentino, L.,
 Zlatav, T., McCombie, W.R., and Martienssen, R.A.,
 Genomic shotgun sequences from Zea mays (methyl-filtered)
 Unpublished (2002)

TITLE

Unpublished (2002)

JOURNAL

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: i181 row: b column: 04

Seq primer: -21M13UnivFwd

Class: shotgun

High quality sequence stop: 569.

Location/Qualifiers

1. 569

/organism="Zea mays"

/mol_type="genomic DNA"

/cultivar="B73"

/db_xref="taxon:4577"

/clone="i181b04"

/lab_host="DH5a"

/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"

/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
 The vector was digested with XbaI and one nucleotide was
 added by fill in in the recessive 3' end. The genomic DNA

was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19, .b/g reads in pUC19). The same ligation was transformed into DH5a."

ORIGIN

Query Match 41.5%; Score 17; DB 28; Length 569;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACTACATAACACACC 23
|||||
DB 56 ACTACATAACACACC 40

RESULT 11

LOCUS CF207448 649 bp mRNA linear EST 01-AUG-2003
DEFINITION CAB20001 IVa_Ra_A09 Cabernet Sauvignon Flower Bloom - CAB2 Vitis vinifera cDNA clone CAB20001 IVa_Ra_A09 3', mRNA sequence.
ACCESSION CF207448
VERSION CF207448.1 GI:33401821
KEYWORDS EST
SOURCE Vitis vinifera
ORGANISM Vitis vinifera

REFERENCE

1 (bases 1 to 649)
Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and Cook, D.
Expressed sequence tags from Vitis vinifera 'Cabernet Sauvignon' berries at various developmental stages
Unpublished (2003)
Contact: Douglas Cook, PhD
CAES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcook@ucdavis.edu
Seq primer: GCCAACGATGCTCTAG.

FEATURES

source

1..649
Location/Qualifiers
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB20001 IVa_Ra_A09"
/sex="Hermaphrodite"
/dev_stage="Bloom"
/clone_lib="Cabernet Sauvignon Flower Bloom - CAB2"
/note="Organ: Flower - Bloom; Vector: pDNR; Site 1: Sfil; Site 2: Sfil; CAB2 is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' Clone 8 berries. Samples were collected at full bloom (80 to 100% flowers showing dehiscence of calypters or caps and anthers fully extended). Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5'-AGCAGTGTATCAACGAGTGGCCATTAGCGCGGG-3' and
5'-ATTCAGAGCGCGGCACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 41.5%; Score 17; DB 14; Length 649;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

6 AACTACATAACACACC 22

DB 132 AACTACATAACACACC 148
|||||

RESULT 12

LOCUS BQ997689 667 bp mRNA linear EST 22-AUG-2002
DEFINITION QGG17C07.YG.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
QGG17C07, mRNA sequence.
ACCESSION BQ997689
VERSION BQ997689.1 GI:22432085
KEYWORDS EST
SOURCE Lactuca sativa
ORGANISM Lactuca sativa

QY 16 CAACCCACTTAGTCAGA 32
|||||

DB 163 CAACCCACTTAGTCAGA 179
|||||

REFERENCE

1 (bases 1 to 667)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik (R.W.Michelmore)
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org (michelmore@vegmail.ucdavis.edu)
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QGG17 row: C column: 07.

JOURNAL

COMMENT

source

1..667
Location/Qualifiers
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGG17C07"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_SEQ=CGAATGCGGG"

FEATURES

source

Query Match 41.5%; Score 17; DB 13; Length 667;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

QY 16 CAACCCACTTAGTCAGA 32
|||||

DB 163 CAACCCACTTAGTCAGA 179
|||||

RESULT 13

LOCUS BQ874693 695 bp mRNA linear EST 15-AUG-2002
DEFINITION QG16A13.YG.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
QG16A13, mRNA sequence.
ACCESSION BQ874693

```

VERSION      BQ874693.1 GI:22261253
KEYWORDS     EST.
SOURCE       Lactuca sativa
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
              asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
              Cichorieae; Lactuca.
REFERENCE    1 (bases 1 to 695)
AUTHORS     Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
              Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
              Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
              Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE       Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
JOURNAL      Unpublished (2002)
COMMENT      Department of Vegetable Crops, R.W.Michelmore Lab
              University of California at Davis (UCD)
              Amundson Hall, UCD, Davis, CA 95616, USA
              Tel: 1-(530)-742-1742
              Fax: 1-(530)-752-9659
              Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
              singleton, see http://cgdb.ucdavis.edu/ for details.
              Plate: QG16 row: A column: 13.
FEATURES     Location/Qualifiers
              source
                1..695
                /organism="Lactuca sativa"
                /mol_type="mRNA"
                /cultivar="Salinas"
                /db_xref="taxon:4236"
                /clone="QG16A13"
                /lab_host="E.coli"
                /clone_lib="QG ABCDI lettuce salinas"
                /note="Vector: pBRCNASTAB; The library was constructed
              from 10 different sources of RNA from a single genotype.
              Separate cDNAs were generated using primers that
              incorporated unique 5' and 3' tags to distinguish each
              source of RNA. cDNAs were then pooled, size-fractionated,
              directionally cloned into a custom medium-copy vector and
              transformations made with four size classes to minimize
              size bias. Details of each source of RNA and library
              construction can be obtained at http://cgdb.ucdavis.edu/
              TAG_TISSUE=roots environmental stress
              TAG_LIB=QG ABCDI lettuce salinas
              TAG_SEQ=ATCTCGGGG"

ORIGIN
Query Match      41.5%; Score 17; DB 13; Length 695;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ATGAACCTACATAACAAC 19
    |||||
Db 205 ATGAACCTACATAACAAC 221

RESULT 14
CG114080/c
LOCUS      CG114080 782 bp DNA linear GSS 20-AUG-2003
DEFINITION PUJGG81TD ZM.0.6.1.0 KB Zea mays genomic clone ZMMBtra0661N17,
            genomic survey sequence.
ACCESSION  CG114080
VERSION     CG114080.1 GI:33997517
KEYWORDS    GSS.
SOURCE      Zea mays
            Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 782)
            Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and

Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUJGG81TB
Contact: Cathy Whitelaw
TIGR Medical Center Drive, Rockville, MD 20850, USA
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
            Location/Qualifiers
            source
              1..782
              /organism="Zea mays"
              /mol_type="genomic DNA"
              /strain="B73"
              /db_xref="taxon:4577"
              /clone="ZMMBtra0661N17"
              /note="Vector: PCR4-TORO; Site 1: EcoRI; 0.6-1.0 kb high
              Cot selected genomic DNA library"

ORIGIN
Query Match      41.5%; Score 17; DB 29; Length 782;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ACTACATAACAACCCACC 23
    |||||
Db 497 ACTACATAACAACCCACC 481

RESULT 15
BX783787
LOCUS      BX783787 870 bp mRNA linear EST 10-DEC-2003
DEFINITION BX783787 XGC-egg Silurana tropicalis cDNA clone Tegg059102 3', mRNA
            sequence.
ACCESSION  BX783787
VERSION     BX783787.1 GI:39690993
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
            Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
            Xenopodinae; Silurana.
            1 (bases 1 to 870)
            Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (11_2003)
            Unpublished (2003)
            Contact: Croning MDR
            Sanger Institute
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: Tegg059102.q1kT7
            Sequencing primer: T7
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
            cDNA was oligo dt primed from sug of poly A+ RNA from egg.
            EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
            5' end and NotI at the 3' end.
            Vector: pCS107; Site 1: EcoRI; Site 2: NotI
            Host: Escherichia coli XL1-blue.
            Location/Qualifiers
            source
              1..870
              /organism="Silurana tropicalis"
              /mol_type="mRNA"
              /db_xref="taxon:8364"
              /clone="Tegg059102"
              /dev_stage="egg"
              /lab_host="Escherichia coli XL1-blue"
              /clone_lib="XGC-egg"

```

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 41.5%; Score 17; DB 13; Length 870;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 ACCTTAGTCAGATACTA 37
|||
Db 287 ACCTTAGTCAGATACTA 303
|||

Search completed: March 25, 2004, 03:06:30
Job time : 2014 secs